

**USPTO RESOURCES ALLOCATED TO SEARCHING 10 SEQUENCES IN 09/394,745**

Lgnth	SEQ ID	DBASE	SRCHTM	JOBTIME	START	FIN	ELPSTM
435	7826	GenEmbl	3842	10519	11:13:08	11:16:33	0:03:25
	7826	GeneSeq	428	5052	11:01:03	11:01:06	0:00:03
	7826	PATS	173	9149	11:42:57	11:43:03	0:00:06
	7826	EST	4942	18145	8:21:05	8:21:08	0:00:03
421	6489	GenEmbl	3842	9912	11:03:09	11:06:26	0:03:17
	6489	GeneSeq	428	4989	11:00:01	11:00:03	0:00:02
	6489	PATS	173	7296	10:51:52	11:12:10	0:20:18
	6489	EST	4942	18128	8:20:48	8:20:51	0:00:03
425	6332	GenEmbl	3842	9715	10:57:42	11:03:09	0:05:27
	6332	GeneSeq	428	4987	10:59:55	11:00:01	0:00:06
	6332	PATS	173	6078	10:51:46	10:51:52	0:00:06
	6332	EST	4942	18125	8:20:45	8:20:48	0:00:03
444	6154	GenEmbl	3842	9388	10:55:05	10:57:42	0:02:37
	6154	GeneSeq	428	4981	10:59:37	10:59:55	0:00:18
	6154	PATS	173	6072	10:51:40	10:51:46	0:00:06
	6154	EST	4942	18122	8:20:41	8:20:45	0:00:04
421	5950	GenEmbl	3842	9231	10:51:11	10:55:05	0:03:54
	5950	GeneSeq	428	4963	10:59:33	10:59:37	0:00:04
	5950	PATS	173	6066	10:51:34	10:51:40	0:00:06
	5950	EST	4942	18118	8:20:37	8:20:41	0:00:04
421	5893	GenEmbl	3842	8997	8:21:14	10:51:11	2:29:57
	5893	GeneSeq	428	4959	9:36:54	10:59:33	1:22:39
	5893	PATS	173	6060	9:10:34	10:51:34	1:41:00
	5893	EST	4942	18114	3:18:43	8:20:37	5:01:54
426	7565	GenEmbl	3842	10314	11:10:50	11:13:08	0:02:18
	7565	GeneSeq	428	5049	11:00:41	11:01:03	0:00:22
	7565	PATS	173	9143	11:22:27	11:42:57	0:20:30
	7565	EST	4942	18142	8:21:01	8:21:05	0:00:04
418	6886	GenEmbl	3842	10176	11:10:46	11:10:50	0:00:04
	6886	GeneSeq	428	5027	11:00:37	11:00:41	0:00:04
	6886	PATS	173	7913	11:22:20	11:22:27	0:00:07
	6886	EST	4942	18138	8:20:58	8:21:01	0:00:03
411	6603	GenEmbl	3942	10172	11:08:54	11:10:46	0:01:52
	6603	GeneSeq	428	5023	11:00:07	11:00:37	0:00:30
	6603	PATS	172	7906	11:12:24	11:22:40	0:10:16
	6603	EST	4942	18135	8:20:54	8:20:58	0:00:04
432	6514	GenEmbl	3842	10060	11:06:26	11:08:54	0:02:28
	6514	GeneSeq	428	4993	11:00:03	11:00:07	0:00:04
	6514	PATS	173	7310	11:12:10	11:12:34	0:00:24
	6514	EST	4942	18131	8:20:51	8:20:54	0:00:03
TTLSECS		93949	402798				
TTLHRS		26.1	111.9			11:54:55	

10:35:30 out of 11:54:55 spent on 5893 just one sequence

09/394745

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 11:13:08 ; Search time 3842.15 Seconds  
(without alignments)  
1867.775 Million cell updates/sec

Title: US-09-394-745-7826  
Perfect score: 435

Sequence: 1 aattcacggccgacgcacg.....cgtcggctttcctgaat 435

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba:  
2: gb\_htg:  
3: gb\_in:  
4: gb\_om:  
5: gb\_ov:  
6: gb\_pat:  
7: gb\_ph:  
8: gb\_pl:  
9: gb\_pr:  
10: gb\_ro:  
11: gb\_sts:  
12: gb\_sy:  
13: gb\_un:  
14: gb\_vi:  
15: em\_ba:  
16: em\_fun:  
17: em\_hum:  
18: em\_in:  
19: em\_om:  
20: em\_or:  
21: em\_ov:  
22: em\_pat:  
23: em\_ph:  
24: em\_pl:  
25: em\_ro:  
26: em\_sts:  
27: em\_sy:

```

28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result		Query				Description	
	No.	Score	Match	Length	DB	ID	
	1	40.8	9.4	4790	8	SBRETROTP2	U07816 Sorghum bic
c	2	39.8	9.1	82453	9	AC004558	AC004558 Homo sapi
c	3	39.4	9.1	193829	9	AC012154	AC012154 Homo sapi
	4	38.6	8.9	72614	9	HS496N17	AL031321 Human DNA
	5	37.8	8.7	73916	9	AP000765	AP000765 Homo sapi
c	6	37.8	8.7	122592	3	CEY60A3A	AL117207 Caenorhab
	7	37.8	8.7	190739	2	AP001361	AP001361 Homo sapi
	8	37.4	8.6	108464	2	AP000710	AP000710 Homo sapi
c	9	37.4	8.6	160759	9	HS297A17	AL513503 Homo sapi
	10	37.4	8.6	176210	2	HS520K3	AL450004 Homo sapi
	11	36.8	8.5	22887	3	CER01H5	Z68007 Caenorhabdi
c	12	36.8	8.5	72356	2	AC084841	AC084841 Homo sapi
	13	36.6	8.4	114084	2	AC009197	AC009197 Drosophil
c	14	36.6	8.4	134580	2	AC025359	AC025359 Homo sapi
	15	36.2	8.3	1141	6	AX083744	AX083744 Sequence
	16	36.2	8.3	145035	9	CNS07EY	AL450442 Human chr
c	17	36.2	8.3	149409	9	AC004081	AC004081 Homo sapi
c	18	36.2	8.3	174232	2	AC026251	AC026251 Homo sapi
c	19	36.2	8.3	178273	2	AC005308	AC005308 Plasmodiu
c	20	36	8.3	140892	2	AC016204	AC016204 Homo sapi
c	21	36	8.3	152409	2	PFMAL1P1	AL031744 Plasmodiu
c	22	36	8.3	186135	2	AC079914	AC079914 Homo sapi
	23	36	8.3	215046	2	AC011767	AC011767 Homo sapi
c	24	35.8	8.2	394	6	AX156156	AX156156 Sequence
	25	35.8	8.2	111489	2	AC084149	AC084149 Homo sapi
c	26	35.8	8.2	129854	33	AC021537	Ac021537 Homo sapi
c	27	35.8	8.2	141016	2	AC092651	AC092651 Homo sapi
	28	35.8	8.2	198146	2	AC074158	AC074158 Mus muscu
c	29	35.4	8.1	122332	2	AC092390	AC092390 Oryza sat
	30	35.4	8.1	222016	2	AC023048	AC023048 Mus muscu
	31	35.2	8.1	2664	9	AF142573	AF142573 Homo sapi
	32	35.2	8.1	2667	9	AF329197	AF329197 Homo sapi
c	33	35.2	8.1	81971	9	AC018753	AC018753 Homo sapi
	34	35.2	8.1	148569	2	AC034292	AC034292 Homo sapi
	35	35.2	8.1	184558	2	AC020570	AC020570 Homo sapi
	36	35.2	8.1	234498	2	AC021077	AC021077 Homo sapi
	37	35	8.0	106730	8	ATF12M12	AL355775 Arabidops
c	38	35	8.0	200087	9	AL354821	AL354821 Human DNA

39	34.8	8.0	64255	2	AC024343	AC024343 Homo sapi
c 40	34.8	8.0	74998	9	AC009423	AC009423 Homo sapi
41	34.8	8.0	77945	2	AC022837	AC022837 Homo sapi
c 42	34.8	8.0	134019	3	AC006471	AC006471 Drosophil
c 43	34.8	8.0	156608	2	AC015512	AC015512 Homo sapi
44	34.8	8.0	169479	9	AC009597	AC009597 Homo sapi
45	34.8	8.0	169600	3	AC092717	AC092717 Drosophil

### ALIGNMENTS

**RESULT** 1  
**SBRETROTP2**  
**LOCUS** SBRETROTP2 4790 bp DNA **PLN** 18-MAR-2000  
**DEFINITION** Sorghum bicolor retrotransposon-like element Levithan, 3' LTR sequence.  
**ACCESSION** U07816  
**VERSION** U07816.1 GI:7262601  
**KEYWORDS** .  
**SEGMENT** 2 of 2  
**SOURCE** sorghum.  
**ORGANISM** Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Sorghum.  
**REFERENCE** 1 (bases 1 to 4790)  
**AUTHORS** Liu,C. and Bennetzen,J.L.  
**TITLE** Characterization of a new family of retrotransposon-like elements in sorghum  
**JOURNAL** Unpublished  
**REFERENCE** 2 (bases 1 to 4790)  
**AUTHORS** Liu,C.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (17-MAR-1994) Chang-Nong Liu, Department of Agronomy,  
Purdue University, West Lafayette, IN 47907, USA  
**FEATURES** Location/Qualifiers  
source 1. .4790  
/organism="Sorghum bicolor"  
/db\_xref="taxon:4558"  
repeat\_region 45. .51  
/rpt\_type=inverted  
LTR 45. .4604  
/note="3' LTR of Levithan, a 15.2 kbp retrotransposon; 4.7  
kpb sequence exists between the 5' and 3' LTR regions, but  
has not been sequenced"  
/label=SRPT1-2  
repeat\_region 4597. .4604  
/rpt\_type=inverted  
**BASE COUNT** 1137 a 1016 c 1138 g 1499 t  
**ORIGIN**

Query Match 9.4%; Score 40.8; DB 8; Length 4790;  
Best Local Similarity 86.5%; Pred. No. 0.66;  
Matches 45; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 218 tccgaatctcgagacgagatttttaagggggagggtgtaacacccag 269  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 6 TTCAAATCTCGGGACGAGATTTGTAAGGAGGGAGGGCTGTAACACCCTAG 57

RESULT 2  
AC004558/c  
LOCUS AC004558 82453 bp DNA PRI 15-APR-1998  
DEFINITION Homo sapiens chromosome 19, overlapping cosmids F20014 and F8998,  
complete sequence.  
ACCESSION AC004558  
VERSION AC004558.1 GI:3047130  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 82453)  
AUTHORS Lamerdin,J.E., McCready,P.M., Skowronski,E., Adamson,A.W.,  
Berkhout-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stilwagen,S.,  
Phan,H., Velasco,N., Garnes,J., Danganan,L., Poundstone,P.,  
Christensen,M., Georgescu,A., Avila,J., Liu,S., Attix,C.,  
Andreise,T., Trankheim,M., Amico-Keller,G., Coefield,J., Duarte,S.,  
Lucas,S., Bruce,R., Thomas,P., Quan,G., Kronmiller,B., Arellano,A.,  
Montgomery,M., Ow,D., Nolan,M., Trong,S., Kobayashi,A., Olsen,A.O.  
and Carrano,A.V.  
TITLE Sequence analysis of a 2.5 Mb region in 19q13.2 containing a  
clustered CEA/PSG gene family  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 82453)  
AUTHORS Lamerdin,J.E.  
TITLE Direct Submission  
JOURNAL Submitted (15-APR-1998) Joint Genome Institute, Lawrence Livermore  
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA  
COMMENT Map and sequence oriented from q centromere to telomere. Accession  
comprised of sequence from cosmid F20014 from bases 1 to 38,269,  
and cosmid F8998 from bases 38,270 to 82,453. No sequence errors  
were detected in overlapping region. Currently there is a small  
sequence gap between cosmid F20014 and F9933 to the left. Cosmid  
F8998 overlaps cosmid F24083 to the right by approx. 4 kb.  
Additional map and sequence information may be obtained at:  
<http://www-bio.llnl.gov/genome/genome.html>.  
FEATURES Location/Qualifiers  
source 1. .82453  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="F20014-F8998"  
/chromosome="19"  
/map="BCKDHA-D19S217"  
/cell\_line="UV5HL9-5B"  
/clone\_lib="LL19NC02 F chromosome 19-specific cosmid  
library"  
/note="Cosmid library constructed at LLNL from flow-sorted  
chromosomes from hybrid UV5HL9-5B, which carries  
chromosome 19 as its only human chromosome"  
repeat\_region complement(751. .1049)  
/rpt\_family="AluY"

repeat\_region 1050. .1178  
/rpt\_family="MIR"  
repeat\_region complement(1959. .2141)  
/rpt\_family="MIR"  
repeat\_region 2253. .2443  
/rpt\_family="MLT1D"  
repeat\_region complement(2867. .3390)  
/rpt\_family="MER57\_internal"  
repeat\_region complement(3559. .3878)  
/rpt\_family="MER65\_internal"  
repeat\_region 3881. .4168  
/rpt\_family="AluY"  
repeat\_region complement(4173. .4665)  
/rpt\_family="MER65\_internal"  
mRNA complement(4182. .19114)  
/gene="CGM6"  
/product="CGM6"  
gene complement(4182. .19114)  
/gene="CGM6"  
/note="carcinoembryonic antigen precursor"  
repeat\_region 4680. .4978  
/rpt\_family="AluSg"  
repeat\_region 4991. .5017  
/rpt\_family="(TA)n"  
repeat\_region complement(5030. .5109)  
/rpt\_family="MER57\_internal"  
repeat\_region 5201. .5356  
/rpt\_family="MER5B"  
CDS complement(join(6316. .6407,11854. .12108,12527. .12805,  
16611. .16970,17835. .17898))  
/gene="CGM6"  
/note="CARCINOEMBRYONIC ANTIGEN CGM6 PRECURSOR  
(NONSPECIFIC CROSS-REACTING ANTIGEN NCA-95) (ANTIGEN CD67)  
(CD66B)"  
/codon\_start=1  
/product="CGM6\_HUMAN"  
/protein\_id="AAC13659.1"  
/db\_xref="GI:3047131"  
/translation="MGPIISAPSCRWRIPWQGLLLTASLFTFWNPPTTAQLTIEAVPSN  
AAEGKEVLLLHVNLQPDPRGYNWYKGETVDANRRIIGYVISNQQITPGPAYSNRETIY  
PNASLLMRNVTRNDTGSYTLQVIKLNLMSEEVTGQFSVHPETPKPSISSNNSNPVEDK  
DAVAFTCEPETQNTTYLWWVNGQSLPVSPRLQLSNGNRITLLSVTRNDVGPYECEIQ  
NPASANFSDPVTLNVLGYGPDAPTISPSDTYYHAGVNLNLSCAASNPPSQYSWSVNGT  
FQQYTQKLFIPNITTKNSGSYACHTNSATGRNRTTVRMITVSDLVQGSSPGLSARA  
TVSIMIGVLARVALI"  
repeat\_region 6759. .7000  
/rpt\_family="LTR16B"  
repeat\_region 7274. .7298  
/rpt\_family="(CA)n"  
repeat\_region complement(7409. .7581)  
/rpt\_family="MIR"  
repeat\_region complement(7618. .7782)  
/rpt\_family="MIR"  
repeat\_region 7783. .8082  
/rpt\_family="L1"  
repeat\_region 8089. .8403  
/rpt\_family="AluJo"

repeat\_region 8423. .8548  
/rpt\_family="L1"  
repeat\_region 8599. .10763  
/rpt\_family="L1MB5"  
repeat\_region 10793. .11049  
/rpt\_family="AluSp"  
misc\_feature complement(11854. .12075)  
/gene="CGM6"  
/note="predicted exon, program: grail2exons\_human\_1.3,  
frame: 0, quality: excellent, score: 100.000"  
misc\_feature complement(12506. .12805)  
/gene="CGM6"  
/note="predicted exon, program: grail2exons\_human\_1.3,  
frame: 2, quality: excellent, score: 90.000"  
repeat\_region 13719. .13945  
/rpt\_family="L1MB1"  
repeat\_region 13951. .14161  
/rpt\_family="MER46"  
repeat\_region 15312. .15360  
/rpt\_family="MIR"  
repeat\_region 15393. .15687  
/rpt\_family="AluSc"  
misc\_feature complement(16603. .16970)  
/gene="CGM6"  
/note="predicted exon, program: grail2exons\_human\_1.3,  
frame: 1, quality: excellent, score: 77.000"  
repeat\_region complement(17047. .17082)  
/rpt\_family="(CA)n"  
repeat\_region 17119. .17429  
/rpt\_family="LINE2"  
misc\_feature complement(17835. .17938)  
/gene="CGM6"  
/note="predicted exon, program: grail2exons\_human\_1.3,  
frame: 1, quality: excellent, score: 75.000"  
repeat\_region 18196. .18267  
/rpt\_family="(CA)n"  
repeat\_region complement(20356. .20573)  
/rpt\_family="AluSg"  
repeat\_region complement(20574. .20893)  
/rpt\_family="L1"  
repeat\_region complement(20912. .22752)  
/rpt\_family="L1PB1"  
repeat\_region 21269. .21327  
/rpt\_family="(CA)n"  
repeat\_region complement(23336. .23488)  
/rpt\_family="MER54"  
repeat\_region complement(23581. .23869)  
/rpt\_family="AluSx"  
repeat\_region complement(24094. .24466)  
/rpt\_family="MLT1A2"  
exon complement(24236. .24326)  
/note="DPS similarity to (X16455) pCEA80-11 protein (647  
AA) [Homo sapiens]."  
/pseudo  
repeat\_region complement(24472. .24753)  
/rpt\_family="AluSx"  
repeat\_region complement(25716. .25842)

```

repeat_region           /rpt_family="L1MB4"
repeat_region           complement(25844. .26146)
repeat_region           /rpt_family="AluSg"
repeat_region           complement(26147. .26680)
repeat_region           /rpt_family="L1MB5"
repeat_region           complement(26722. .27449)
repeat_region           /rpt_family="L1MB5"
misc_feature            27686. .27900
                           /note="predicted exon, program: grail2exons_human_1.3,
                           frame: 1, quality: good, score: 66.000"
misc_feature            28125. .28216
                           /note="predicted exon, program: grail2exons_human_1.3,
                           frame: 2, quality: excellent, score: 82.000"
repeat_region           complement(29034. .29307)
repeat_region           /rpt_family="AluSg"
repeat_region           29574. .29695
repeat_region           /rpt_family="MER20"
repeat_region           complement(30648. .30860)
repeat_region           /rpt_family="L1"
repeat_region           complement(30935. .31675)
repeat_region           /rpt_family="L1ME3"
repeat_region           31983. .32149
repeat_region           /rpt_family="AluJb"
repeat_region           32551. .32694
repeat_region           /rpt_family="MER4D"
repeat_region           32749. .34381
repeat_region           /rpt_family="SVA"
repeat_region           34593. .34758
repeat_region           /rpt_family="MER57_internal"
repeat_region           34825. .35547
repeat_region           /rpt_family="MER4D"
repeat_region           35663. .35731
repeat_region           /rpt_family="MER57_internal"
repeat_region           36428. .36624
repeat_region           /rpt_family="MER57_internal"
repeat_region           36633. .37326

```

```

Query Match      9.1%;  Score 39.8;  DB 9;  Length 82453;
Best Local Similarity 51.4%;  Pred. No. 1.8;
Matches 92;  Conservative 0;  Mismatches 87;  Indels 0;  Gaps 0;

```

```

Qy      65 tgggtggacatctctaaatttagcttaaggcgatacatgttatgtccactagagaaaacaaca 124
        ||| ||| |||||| | ||| |||| | | | | | | | | | | | | | | | | | | | |
Db     10061 TGTTGAGCATCTTGCATGTGCTCATTGGCCATTGTATATCTTCCTGGAGAAATCTCT 10002

```

```

Qy      125 tcctgagacactcactttattggaaatgtctcgcgattatcgctgatgtggacatgtg 184
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     10001 ATTCAGTCTTGTCCTTTAGTTGGGTTTGAGATTGCTGTTGGATTGTA 9942

```

```

Qy      185 ttacatgcttctactctaaaaagtctttgctccaatctcgagacgagattttt 243
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     9941 GTAGTTCTCATATACTCTGAAAATTGATCCCTTATCATACATGATTACAAATTTT 9883

```

```

RESULT    3
AC012154/c
LOCUS      AC012154      193829 bp      DNA          PRI          28-JUL-2001

```

DEFINITION Homo sapiens 3 BAC RP11-48H24 (Roswell Park Cancer Institute Human BAC Library) complete sequence.

ACCESSION AC012154

VERSION AC012154.16 GI:14578093

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 193829)

AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhajj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., He, X., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Moore, S., Morgan, M., Moorish, T., Morris, S., Moser, M., Neal, D., Nelson, D., Newtson, J., Newtson, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telfrod, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Naylor, S.L., Weinstock, G. and Gibbs, R.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 193829)

AUTHORS Worley, K.C.

TITLE Direct Submission

JOURNAL Submitted (21-OCT-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

REFERENCE      Baylor Plaza, Houston, TX 77030, USA  
AUTHORS      3 (bases 1 to 193829)  
TITLE      Worley,K.C.  
JOURNAL      Direct Submission  
Submitted (07-JUL-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

REFERENCE      4 (bases 1 to 193829)  
AUTHORS      Worley,K.C.  
TITLE      Direct Submission  
JOURNAL      Submitted (10-JUL-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

REFERENCE      5 (bases 1 to 193829)  
AUTHORS      Worley,K.C.  
TITLE      Direct Submission  
JOURNAL      Submitted (28-JUL-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

COMMENT      On Jun 30, 2001 this sequence version replaced gi:14547736.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES	Location/Qualifiers
source	1. .193829 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="3" /clone="RP11-48H24"
misc_feature	1. .2008 /note="Overlaps bases 152829. .154828 of AC069070" /function="Overlaps with adjacent clone AC069070"
repeat_region	244. .365 /rpt_family="AluJb"
repeat_region	366. .412 /rpt_family="(CA)n"
repeat_region	413. .562 /rpt_family="AluJb"
repeat_region	563. .584 /rpt_family="(TAAA)n"
repeat_region	607. .787 /rpt_family="MER20"
repeat_region	complement(788. .862) /rpt_family="MER2"
repeat_region	complement(901. .1072) /rpt_family="MER2"
repeat_region	1684. .1896 /rpt_family="HAL1"
repeat_region	complement(1897. .2204) /rpt_family="AluSg1"
repeat_region	2205. .2228 /rpt_family="HAL1"
repeat_region	2229. .2251 /rpt_family="(CA)n"
repeat_region	2252. .2345 /rpt_family="HAL1"
repeat_region	2521. .2646 /rpt_family="MLT1I"
repeat_region	3005. .3126 /rpt_family="HAL1"
repeat_region	3181. .3228 /rpt_family="(TATG)n"
repeat_region	3230. .3273 /rpt_family="AT_rich"
repeat_region	3503. .3628 /rpt_family="LTR12"
repeat_region	3629. .4072 /rpt_family="LTR12"
STS	4600. .4699 /standard_name="76067"
repeat_region	complement(7298. .7577) /rpt_family="AluJo"
repeat_region	8422. .8458 /rpt_family="AT_rich"
repeat_region	complement(8696. .8873) /rpt_family="MLT1D"
repeat_region	9120. .9187 /rpt_family="(CATATA)n"
repeat_region	9695. .10135

```
repeat_region /rpt_family="MLT1C"
repeat_region 10323. .10355
repeat_region /rpt_family="(TG)n"
repeat_region 10355. .10398
repeat_region /rpt_family="(GA)n"
repeat_region complement(10948. .11063)
repeat_region /rpt_family="L1P4"
repeat_region 11038. .11356
repeat_region /rpt_family="L1PA12"
repeat_region 12326. .12366
repeat_region /rpt_family="Harlequin"
repeat_region 13048. .13086
repeat_region /rpt_family="GA-rich"
repeat_region 13117. .13360
repeat_region /rpt_family="GA-rich"
repeat_region 14717. .14744
repeat_region /rpt_family="(TA)n"
repeat_region 14745. .14778
repeat_region /rpt_family="MADE1"
repeat_region 14779. .14822
repeat_region /rpt_family="(TA)n"
repeat_region complement(14871. .15334)
repeat_region /rpt_family="MLT1E2"
repeat_region 16227. .16286
repeat_region /rpt_family="(TA)n"
repeat_region 16525. .16573
repeat_region /rpt_family="AT_rich"
repeat_region 17085. .17176
```

Query Match 9.1%; Score 39.4; DB 9; Length 193829;  
Best Local Similarity 48.1%; Pred. No. 2.6;  
Matches 112; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

```

Qy      181 tgtgttacatgcttctactcttaaaagtctttgctccgaatctcgagacgagattat 240
       ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Pb      65065 TGTGTTATTTATTCAATATTGAGAAAATTGAGAATTTCCTTATTGGCCTATCTGAT 65006

```

```

Qy      241 tttaaggggggagggctgtacaccccccaggtttatattctgctcgacaacgagttatgg 300
       ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Pb      65005 TTCTACTCTGGACAGAACCTTATCTATCTAAATCGACACACACTCACTTAATG 64946

```

```

Qy      301 aattaaggcacgttatatacgtgaatgaaacagataactaaaatttaatcatttcgtatc 360
       |||||           |||||||| | | | | | | | | | | | | | | | | | | | |
Db     64945 CATTACATTTTATATCTGTAATGGAATTGAAATCTTAGACTAATTCCCTACCACATTT 64886

```

RESULT 4  
HS496N17  
LOCUS HS496N17 72614 bp DNA PRI 23-NOV-1999  
DEFINITION Human DNA sequence from clone 496N17 on chromosome 6p11.2-12.3  
Contains EST, GSS, complete sequence.  
ACCESSION AL031321  
VERSION AL031321.1 GI:3676209

KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 72614)  
 AUTHORS Phillips, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (15-OCT-1998) Sanger Centre, Hinxton, Cambridgeshire,  
           CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone  
           requests: clonerequest@sanger.ac.uk  
 COMMENT On Sep 30, 1998 this sequence version replaced gi:3550750.  
         During sequence assembly data is compared from overlapping clones.  
         Where differences are found these are annotated as variations  
         together with a note of the overlapping clone name. Note that the  
         variation annotation may not be found in the sequence submission  
         corresponding to the overlapping clone, as we submit sequences with  
         only a small overlap as described above.  
         This sequence is the entire insert of clone 496N17. This sequence  
         has been finished according to sequence map criteria as follows. An  
         attempt is made to resolve all sequencing problems, such as  
         compressions and repeats, but not necessarily within known  
         annotated human repeat sequence elements (e.g. Alu). Where the  
         sequence is ambiguous, there is an annotation using the 'unsure'  
         feature key.  
         This sequence was generated from part of bacterial clone contigs of  
         human chromosome 6, constructed by the Sanger Centre Chromosome 6  
         Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr6>  
         496N17 is from the library RPCI3 constructed at the Roswell Park  
         Cancer Institute by the group of Pieter de Jong. For further  
         details see <http://bacpac.med.buffalo.edu/> VECTOR: pCYPAC2.  
 FEATURES  
 source                 Location/Qualifiers  
           1. .72614  
           /organism="Homo sapiens"  
           /db\_xref="taxon:9606"  
           /chromosome="6"  
           /map="p11.2-12.3"  
           /clone="RP3-496N17"  
           /clone\_lib="RPCI-3"  
 prim\_transcript <635. .>1795  
           /note="match: multiple ESTs; match: C16189 AA113351  
           AA004642 AA767319 AI016738 AA253071 AA298736 AA279359  
           AA372233 AA256610 AA908947 AA280336 W24074 AI042215  
           AA134563 C16263 AI097375 AI095854 AA938204 AI085203  
           AA954988 AA779659 AA418951 AA418744 AA280378 N99001  
           AA113433 AA136041 AA488507 AA937184 AI018267 AA993960  
           AI018739 AA134562 D11563 AI080441 C16239 D57063 AA436556  
           AA741485 AA436583"  
 repeat\_region 1820. .1972  
           /note="L1M3 repeat: matches 5612. .5773 of consensus"  
 repeat\_region 2963. .3013  
           /note="Alu repeat: matches 250. .300 of consensus"  
 repeat\_region 3180. .3473  
           /note="AluSg repeat: matches 1. .293 of consensus"  
 prim\_transcript <3851. .>4333  
           /note="match: multiple ESTs; match: AA233335 AA029031

AA490697 AI093515 AA233336 AA028109 AA406564 AA034020"  
misc\_feature 4042. .4536  
/note="match: GSS B81642 clone R-14D8"  
prim\_transcript <4374. .>7149  
/note="match: multiple ESTs; match: R07340 AA150152 C01711  
AA705004 AA897243 AA704987 AA134570 AA694297 D79493 W61217  
AA605713 R11596 AI038012 AI078466 AA788639 R37178 R07388  
AA605762 AA490474 AA134569 R19387 AA137200 AA779313  
AA137199 D61796 AA705300 D61863 D62821 D61963 N39560  
D79490 T16460 AA605640 N77551 R24690 D79953 D79395 R35591  
AA605615 AA232916 AA033539 AA917689 AI095486 D61830 W60970  
T23992 W60745 W60718 D79302 AA028110 AA411698"  
misc\_feature 7088. .7526  
/note="match: GSS B38883"  
repeat\_region 7570. .7986  
/note="L1MB3 repeat: matches 5745. .6176 of consensus"  
repeat\_region 8306. .8606  
/note="AluSx repeat: matches 2. .302 of consensus"  
repeat\_region 8724. .9001  
/note="AluJo repeat: matches 1. .282 of consensus"  
repeat\_region 9203. .9320  
/note="L2 repeat: matches 2576. .2694 of consensus"  
repeat\_region 9466. .9598  
/note="AluJ repeat: matches 1. .133 of consensus"  
repeat\_region 12285. .12764  
/note="MLT1F repeat: matches 1. .541 of consensus"  
repeat\_region 13596. .13709  
/note="2 copies 57 mer 97% conserved"  
repeat\_region 13883. .13964  
/note="2 copies 41 mer 92% conserved"  
repeat\_region 14593. .14933  
/note="AluJo repeat: matches 1. .302 of consensus"  
repeat\_region 15146. .15389  
/note="AluY repeat: matches 56. .298 of consensus"  
repeat\_region 15392. .15553  
/note="AluJ repeat: matches 136. .297 of consensus"  
repeat\_region 16173. .16301  
/note="LTR16C repeat: matches 195. .329 of consensus"  
repeat\_region 17115. .17419  
/note="AluSg repeat: matches 1. .305 of consensus"  
repeat\_region 17780. .17923  
/note="MER34 repeat: matches 401. .539 of consensus"  
repeat\_region 17932. .18204  
/note="AluSx repeat: matches 37. .310 of consensus"  
repeat\_region 18210. .18618  
/note="MER39 repeat: matches 6. .414 of consensus"  
repeat\_region 19859. .20163  
/note="AluSx repeat: matches 1. .306 of consensus"  
repeat\_region 21790. .21963  
/note="87 copies 2 mer at 76% conserved"  
repeat\_region 21976. .22025  
/note="inverted 79% conserved, 48bp loop"  
repeat\_region complement(22074. .22122)  
/note="inverted 79% conserved, 48bp loop"  
repeat\_region 22689. .23309  
/note="L2 repeat: matches 241. .899 of consensus"  
repeat\_region 24719. .24796

repeat\_region /note="MER47 repeat: matches 2224. .2301 of consensus"  
24838. .24907  
repeat\_region /note="MER47 repeat: matches 2251. .2320 of consensus"  
25224. .25507  
repeat\_region /note="AluSx repeat: matches 2. .305 of consensus"  
25571. .26216  
repeat\_region /note="MER82 repeat: matches 2. .653 of consensus"  
26827. .26926  
repeat\_region /note="50 copies 2 mer aa 64% conserved"  
27599. .27726  
repeat\_region /note="MIR repeat: matches 68. .221 of consensus"  
27749. .28026  
repeat\_region /note="MER46B repeat: matches 4. .276 of consensus"  
28027. .28347  
repeat\_region /note="AluJo repeat: matches 1. .310 of consensus"  
28828. .28973  
repeat\_region /note="MER5A repeat: matches 39. .187 of consensus"  
29133. .29334  
repeat\_region /note="LTR16B repeat: matches 209. .451 of consensus"  
29400. .29457  
repeat\_region /note="MER5B repeat: matches 112. .177 of consensus"  
29458. .29765  
repeat\_region /note="AluJo repeat: matches 1. .306 of consensus"  
29766. .29821  
repeat\_region /note="MER5B repeat: matches 84. .112 of consensus"  
29822. .30180  
repeat\_region /note="MER1B repeat: matches 1. .337 of consensus"  
30181. .30206  
repeat\_region /note="MER5B repeat: matches 56. .84 of consensus"  
30444. .31174  
repeat\_region /note="L1MD repeat: matches -2. .681 of consensus"  
31175. .31461  
repeat\_region /note="AluJb repeat: matches 1. .298 of consensus"  
31462. .32825  
repeat\_region /note="L1MD repeat: matches 681. .2286 of consensus"  
32826. .33156  
repeat\_region /note="MER7A repeat: matches 1. .345 of consensus"  
33157. .33252  
repeat\_region /note="L1MD repeat: matches 2286. .2374 of consensus"  
33236. .33366  
repeat\_region /note="L1MD2 repeat: matches 2600. .2753 of consensus"  
33377. .33904  
repeat\_region /note="MER41B repeat: matches 114. .597 of consensus"  
33923. .34163  
repeat\_region /note="L1MD2 repeat: matches 2770. .3002 of consensus"  
34164. .34465  
repeat\_region /note="AluSx repeat: matches 1. .306 of consensus"  
34466. .35427  
repeat\_region /note="L1MD2 repeat: matches 3002. .3933 of consensus"  
35428. .35736  
repeat\_region /note="AluSx repeat: matches 1. .307 of consensus"  
35737. .37709  
repeat\_region /note="L1MD2 repeat: matches 3933. .6142 of consensus"  
39635. .40327  
repeat\_region /note="L1ME3A repeat: matches 5460. .6152 of consensus"  
40481. .41013  
repeat\_region /note="L1MB3 repeat: matches 5297. .5822 of consensus"

```

repeat_region    41014. .41128
/repeat: matches 2. .125 of consensus"
repeat_region    41129. .41481
/repeat: matches 5822. .6176 of consensus"
repeat_region    41632. .42175
/repeat: matches 6. .545 of consensus"
repeat_region    42835. .42964
/repeat: matches 99. .247 of consensus"
repeat_region    42927. .43739
/L2 repeat: matches 1873. .2712 of consensus"
repeat_region    44220. .44527
/AluSq repeat: matches 1. .311 of consensus"
repeat_region    45541. .45758
/MIR repeat: matches 23. .255 of consensus"
repeat_region    45980. .46060
/L1MEc repeat: matches 1496. .1577 of consensus"
repeat_region    46336. .46647
/AluSg repeat: matches 1. .308 of consensus"
repeat_region    47899. .47986
/MER20 repeat: matches 36. .126 of consensus"
repeat_region    48260. .48652
/L1M4 repeat: matches 5194. .5636 of consensus"
misc_feature    <48857. .49248
/match: GSS AQ196935 clone 2375L23"

```

Query Match 8.9%; Score 38.6; DB 9; Length 72614;  
Best Local Similarity 48.0%; Pred. No. 3.9;  
Matches 110; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

Qy	187 acatgcttctactctaaaagtctttgcgtccgaatctcgagacgagattatTTAAG	246
Db	51157 AAAAGGGTATCTACTCCAAGTTGTGTTATTTCAAATTTCATACTAGCTCATTAGTA	51216
Qy	247 gggggaggggctgtaacacccccagggtttatattctgctcgacaacgagtatgaaattaa	306
Db	51217 TTGTCATTGCACTAGCATCACTGTTGGTATTCTCCAGCGTGTGAGATATTAAATTAACTC	51276
Qy	307 gcacgttatatcagtaatgaaacagataactaaaattaatcatttcgctatcgcgatt	366
Db	51277 CTGTTAGAAAAGTTGAAGATAACAGCTACTAACTGAGGATTAAATTATGGCTTA	51336
Qy	367 tttatatcgtatctgttccatctgtcggtgacatcattttatt	415
Db	51337 ATAATAATTAGTAACTTGCTGCTGTGTTATTATTATT	51385

RESULT 5  
AP000765  
LOCUS AP000765 73916 bp DNA PRI 18-JUL-2001  
DEFINITION Homo sapiens genomic DNA, chromosome 11q, clone:RP11-816P15,  
complete sequence.  
ACCESSION AP000765  
VERSION AP000765.5 GI:14861099  
KEYWORDS HTG.  
SOURCE Homo sapiens DNA, clone:RP11-816P15.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 AUTHORS 1 (bases 1 to 73916)  
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
 TITLE Homo sapiens genomic DNA  
 JOURNAL Published Only in Database (1999) In press  
 REFERENCE 2 (bases 1 to 73916)  
 AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-NOV-1999) Masahira Hattori, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,  
 Tel:81-45-503-9111, Fax:81-45-503-9170)  
 COMMENT On Jul 17, 2001 this sequence version replaced gi:11994960.  
 FEATURES Location/Qualifiers  
 source 1. .73916  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="11"  
 /map="11q"  
 /clone="RP11-816P15"  
 BASE COUNT 24499 a 15941 c 15070 g 18406 t  
 ORIGIN

Query Match 8.7%; Score 37.8; DB 9; Length 73916;  
 Best Local Similarity 55.8%; Pred. No. 6.5;  
 Matches 72; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 68 tggacatctctaaattagcttaaggcgatacatgttatgtccactagagaaaacaacatcc 127  
 ||||||| ||||| |||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 Db 58487 TGGACATATCTAATTAAAATAACAATATCCATTTGTATATGCAATAAACCTACCTTC 58546

Qy 128 tgagacactcaccttatttggaaatgtctcgcgattatcgctgatgtggacatgttta 187  
 ||| ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 Db 58547 TGTAAGATAACAATTCTGGTCTGTTGCTGGATAACCCTAAATGAGGTTTTTT 58606

Qy 188 catgcttct 196  
 | | | |  
 Db 58607 TTTAATTAT 58615

RESULT 6  
 CEY60A3A/c  
 LOCUS CEY60A3A 122592 bp DNA INV 20-JUN-2001  
 DEFINITION Caenorhabditis elegans cosmid Y60A3A, complete sequence.  
 ACCESSION AL117207 AL021574  
 VERSION AL117207.1 GI:5832916  
 KEYWORDS HTG.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 REFERENCE 1 (bases 1 to 122592)  
 AUTHORS none.

TITLE Genome sequence of the nematode *C. elegans*: a platform for investigating biology. The *C. elegans* Sequencing Consortium  
 JOURNAL Science 282 (5396), 2012-2018 (1998)  
 MEDLINE 99069613  
 REMARK The *C. elegans* Sequencing Consortium.  
 REFERENCE 2 (bases 1 to 122592)  
 AUTHORS Williams, L.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-APR-1999) Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1RQ, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rw@nematode.wustl.edu  
 COMMENT On May 14, 2001 this sequence version replaced gi:4914474.  
 Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.  
 Current sequence finishing criteria for the *C. elegans* genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.  
 IMPORTANT: This sequence is not the entire insert of clone Y60A3A. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.  
 The true left end of clone Y60A3 is at 1 in this sequence. The true left end of clone Y102G3 is at 3623 in this sequence. The true left end of clone Y113G7 is at 47717 in this sequence. The true right end of clone Y116F11 is at 47722 in this sequence. The true right end of clone Y60A3 is at 122592 in this sequence. The start of this sequence (1..108) overlaps with the end of sequence AL132943. The end of this sequence (122480..122592) overlaps with the start of sequence AL132858.  
 For a graphical representation of this sequence and its analysis see:- <http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=Y60A3A>.

FEATURES Location/Qualifiers  
 source 1..122592  
   /organism="Caenorhabditis elegans"  
   /db\_xref="taxon:6239"  
   /chromosome="V"  
   /clone="Y60A3A"  
 gene complement(join(5191..5202,5898..6056,6642..6721,  
   6963..7060,9809..9933,10011..10097,10147..10203,  
   10257..10376,10728..10835,11185..11321,11428..11566,  
   11610..11740,11797..11933,11987..12027))  
   /gene="Y60A3A.12"  
 CDS complement(join(5191..5202,5898..6056,6642..6721,  
   6963..7060,9809..9933,10011..10097,10147..10203,  
   10257..10376,10728..10835,11185..11321,11428..11566,  
   11610..11740,11797..11933,11987..12027))  
   /gene="Y60A3A.12"  
   /note="contains similarity to Pfam domain: PF00069 (Eukaryotic protein kinase domain), Score=234.0, E-value=6.7e-67, N=2; PF00498 (FHA domain), Score=36.0, E-value=2.7e-07, N=1  
   cDNA EST yk523b3.3 comes from this gene

cDNA EST yk523b3.5 comes from this gene  
cDNA EST EMBL:AB049441 comes from this gene  
cDNA EST AB041996 comes from this gene"  
/codon\_start=1  
/protein\_id="CAB60407.2"  
/db\_xref="GI:14530652"  
/db\_xref="SPTREMBL:Q9U1Y5"  
/translation="MVRGTKRRRSSAEKPIVVVPVTRDDTMPVDEDLVGESQCAASK  
PFAKLGVRRGISSIDLADDHFVCGRGSDAPTNFNFNSQVAKDVGLYRFISKIQFSID,  
RDTETRRIYLHDHSRNGTLVNQEMIGKGLSRELMNGDLISIGIPALIIFVYESADADH  
HPEELTKKYHVTSHSLGGFGKVLLGYKKSDRSVVAIKQLNTQFSTRCSRAIAKTRD  
IRNEVEVMKKLSHPNIVAIYDWITVAKYSYMVIEYVGGEFFSKVVDSKYNRMGLGES  
LGKYFAFQLIDAILYLHSVGVICHEDIK PENILCSDKAERCILKLTDFGMAKNSVRMK  
TRCGTPSYNAPEIVANEVGEYTPKVDIWSLGCVLFITFSGYPPFSEETDMTMDEQVL  
TGRLIFHAQWRRITVETQNMIKWMLTVEPSNRPSAVELMSTQWMKCADCRTAKQDILK  
SIKPISAAAPAALQTQAGPVKKAKM"  
join(12623. .12737,12807. .12957,14005. .14176,15519.  
gene  
.15682,  
15810. .15975)  
/gene="Y60A3A.19"  
join(12623. .12737,12807. .12957,14005. .14176,15519.  
CDS  
.15682,  
15810. .15975)  
/gene="Y60A3A.19"  
/note="cDNA EST yk391g9.3 comes from this gene  
cDNA EST yk417g4.3 comes from this gene  
cDNA EST yk391g9.5 comes from this gene  
cDNA EST yk417g4.5 comes from this gene"  
/codon\_start=1  
/protein\_id="CAB60403.2"  
/db\_xref="GI:9367164"  
/db\_xref="SPTREMBL:Q9U1Y8"  
/translation="MDYPGTGVKQPEVSIDLNGFYSSNFEQFEHPDQSSNAPNSGSIS  
SPTGQISAQAYRTNAGAAGKFMENSGFGWLLEVNEEDSDQIPLLEELDIDLTDIYYK  
IRCVLLPLPYFRMKNLNVRESPDFWGPLAVVLAFAILSLYQFGVVSVIITIWFCCGF  
MVYFIARALGGDVGVYSQLVGLGIVGYCLIPLVVTSЛИTPLFSSFRLLSNGLGMFTIWSV  
YSAGTLLCVDELQAKKPLVVYPVFLYIYFYSLYSGV"  
complement(join(17723. .18014,18081. .18181,18227. .18344,  
18452. .18648))  
/gene="Y60A3A.16"  
complement(join(17723. .18014,18081. .18181,18227. .18344,  
18452. .18648))  
/gene="Y60A3A.16"  
/note="contains similarity to Pfam domain: PF02149 (Kinase  
associated domain 1), Score=73.0, E-value=2.1e-18, N=1"  
/codon\_start=1  
/protein\_id="CAB60401.1"  
/db\_xref="GI:6425368"  
/db\_xref="SPTREMBL:Q9U1Z0"  
/translation="MMEPIEECDENVGAFQWQQASETLAIRQRGFALETTDQEDMP  
RAVRFTWNLKKTSMLEPDEILKEIQKVLGSYRIDYEQQKRFLLRCSSHVDPLTDASVKW  
EIEVCTLPRLYLNGVHFQRISGSSDFKNIITKISEEDTFALEAQKTRKNKESLILKV  
LVLDPLAIILLPHYRNKLFGIAKKHIFPIFPMLSAEFLSVFDTSSFNLYFKCFIYS  
IVILPSNFSIARFFHCF"  
join(20307. .20369,20427. .20672,20745. .20915)  
/gene="Y60A3A.18"  
join(20307. .20369,20427. .20672,20745. .20915)  
CDS

/gene="Y60A3A.18"  
/note="contains similarity to Pfam domain: PF01466 (Skp1 family), Score=167.6, E-value=6.8e-47, N=1  
cDNA EST yk514e10.5 comes from this gene  
cDNA EST yk514e10.3 comes from this gene"  
/codon\_start=1  
/protein\_id="CAB60402.1"  
/db\_xref="GI:6425369"  
/db\_xref="SPTRREMBL:Q9U1Y9"  
/translation="MSDADSQKQIKLISDDKTFTVSRKVVISQSKTISGFTSEDTIPL  
PKVTSAILEKIITWCEHHADDEPKKVEKIEKGNNKKTVEISEWDAEFMKV  
DQGTLFIEII  
LAANYLDIRGLLEVTTQNVANMMKGKTPSQVRTLKFIDNFSEELEAMKKGN  
AWCED"  
gene complement(join(21462. .21661,21699. .21882,21932. .22021,  
22077. .22207,22706. .22859))  
/gene="Y60A3A.15"  
CDS complement(join(21462. .21661,21699. .21882,21932. .22021,  
22077. .22207,22706. .22859))  
/gene="Y60A3A.15"  
/note="contains similarity to Pfam domain: PF00104  
(Ligand-binding domain of nuclear hormone receptor),  
Score=-4.7, E-value=0.011, N=1"  
/codon\_start=1  
/protein\_id="CAB60400.1"  
/db\_xref="GI:6425367"  
/db\_xref="SPTRREMBL:Q9U1Z1"  
/translation="MLSAEFLSVFDTSSFNLYFKCFIYSIVILPSNFSIARFFHCFW  
LFWVLQLKIQKFDSYTVEYVELPTFQEYSKALLPAHSDGISFQHGH  
VEQEIDTSSVLA  
HLKTALQWVQQFSLFAVLSDVEKSQIILTKWPHLLCLALFENA  
EKIFIDEKFAQLAEK  
FNVLKVSACQDYFLLKGIMIFTESQFNTNNGA  
DLKFDRQLDICIGLLNQLHSESSSKSKS  
GRLIFLLGELKSYS  
TROLESLLDKTCEIVISFL"  
gene complement(join(23935. .24106,24773. .25104,26702. .26883,  
26998. .27150,28664. .28913))  
/gene="Y60A3A.14"  
CDS complement(join(23935. .24106,24773. .25104,26702. .26883,  
26998. .27150,28664. .28913))  
/gene="Y60A3A.14"  
/note="contains similarity to Pfam domain: PF00953  
(Glycosyl transferase), Score=203.7, E-value=9.3e-58, N=1"  
/codon\_start=1  
/protein\_id="CAB60399.1"  
/db\_xref="GI:6425366"  
/db\_xref="SPTRREMBL:Q9U1Z2"  
/translation="MGVICAAYLIVMFMFIPFPFLEWKQGSEFPYEKLLALLSGLIS  
ISTA  
IILGFADDMLDLKWRHKLLFPTLSSLPLLMVYYVSGN  
STTVIVPTIVRHLVQPI  
VLLPVTINISFIYYIFMGMVIVFC  
TNAINILAGINGLESQ  
GQSLVISASVCL  
CNFVQIF  
RF  
SAENSTGF  
WHHTISLYFLLPFTACTAILFY  
FNK  
YPSRVFG  
DTFCYWS  
GMTLAVVS  
ILGHFS  
KTLMLFFV  
PQIINF  
FLY  
SIPQLF  
HLVPC  
PRHRLPKYDP  
KTD  
TVSMSIA  
EFKKT  
DLK  
RLGAL  
FIAV  
CKSIGMLH  
VKE  
VEKDGE  
IYLQ  
INNLTI  
INLV  
LK  
FAGPL  
HEKT  
LNDV  
LMSI  
QILCS  
LLAFF  
FIR  
FYLAS  
LFYDV  
VE"  
gene complement(join(38168. .38545,41313. .41517,42058. .42119))  
/gene="Y60A3A.10"  
CDS complement(join(38168. .38545,41313. .41517,42058. .42119))  
/gene="Y60A3A.10"  
/note="contains similarity to Pfam domain: PF00106 (short  
chain dehydrogenase), Score=13.3, E-value=5.8e-07, N=1"  
/codon\_start=1  
/protein\_id="CAB60405.1"



TITLE           Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
JOURNAL        Direct Submission  
                Submitted (06-MAR-2000) Masahira Hattori, The Institute of Physical  
                and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
                Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,  
                Japan (E-mail:hattori@gsc.riken.go.jp,  
                URL:<http://hgp.gsc.riken.go.jp/>, Tel:81-42-778-9923,  
                Fax:81-42-778-9924)  
COMMENT        On May 30, 2000 this sequence version replaced gi:7209907.  
----- Genome Center  
                Center: RIKEN Genomic Sciences Center(GSC)  
                Center code: RIKEN  
                Web site: <http://hgp.gsc.riken.go.jp/>  
                Contact: hattori@gsc.riken.go.jp  
----- Project Information  
                Center project name: HumDraft11  
                Center clone name: RP11-853020  
----- Summary Statistics  
                Sequencing vector: PCR products; 100% of reads  
                Chemistry: Dye-terminator ET-amersham; 100% of reads  
                Assembly program: Phrap; version 0.990329  
                Consensus quality: 170786 bases at least Q40  
                Consensus quality: 180477 bases at least Q30  
                Consensus quality: 184864 bases at least Q20  
                Insert size: 187739; sum-of-contigs  
                Quality coverage: 4.30x in Q20 bases; sum-of-contigs  
-----  
NOTE: This is a 'working draft' sequence. It currently consists of  
31 contigs. The true order of the pieces is not known and their  
order in this sequence record is arbitrary. Gaps between the  
contigs are represented as runs N, but the exact sizes of the gaps  
are unknown. This record will be updated with the finished sequence  
as soon as it is available and the accession number will be  
preserved  

1	19670	contig of	19670 bp in length
19771	32416	contig of	12646 bp in length
32517	44331	contig of	11815 bp in length
44432	57404	contig of	12973 bp in length
57505	67680	contig of	10176 bp in length
67781	78637	contig of	10857 bp in length
78738	87610	contig of	8873 bp in length
87711	98619	contig of	10909 bp in length
98720	108219	contig of	9500 bp in length
108320	116259	contig of	7940 bp in length
116360	123262	contig of	6903 bp in length
123363	128558	contig of	5196 bp in length
128659	134632	contig of	5974 bp in length
134733	140461	contig of	5729 bp in length
140562	146381	contig of	5820 bp in length
146482	151457	contig of	4976 bp in length
151558	153639	contig of	2082 bp in length
153740	157220	contig of	3481 bp in length
157321	161091	contig of	3771 bp in length
161192	164668	contig of	3477 bp in length
164769	168197	contig of	3429 bp in length
168298	171770	contig of	3473 bp in length
171871	175319	contig of	3449 bp in length

175420	178136 contig of	2717 bp in length
178237	181402 contig of	3166 bp in length
181503	182931 contig of	1429 bp in length
183032	184947 contig of	1916 bp in length
185048	186820 contig of	1773 bp in length
186921	188144 contig of	1224 bp in length
188245	189483 contig of	1239 bp in length
189584	190739 contig of	1156 bp in length

Sequence updated (26-May-2000).

- \* NOTE: This is a 'working draft' sequence. It currently
- \* consists of 31 contigs. The true order of the pieces
- \* is not known and their order in this sequence record is
- \* arbitrary. Gaps between the contigs are represented as
- \* runs of N, but the exact sizes of the gaps are unknown.
- \* This record will be updated with the finished sequence
- \* as soon as it is available and the accession number will
- \* be preserved.
- \*       1     19670: contig of 19670 bp in length
- \*       19671 19770: gap of      100 bp
- \*       19771    32416: contig of 12646 bp in length
- \*       32417 32516: gap of      100 bp
- \*       32517    44331: contig of 11815 bp in length
- \*       44332 44431: gap of      100 bp
- \*       44432    57404: contig of 12973 bp in length
- \*       57405 57504: gap of      100 bp
- \*       57505    67680: contig of 10176 bp in length
- \*       67681 67780: gap of      100 bp
- \*       67781    78637: contig of 10857 bp in length
- \*       78638 78737: gap of      100 bp
- \*       78738    87610: contig of 8873 bp in length
- \*       87611 87710: gap of      100 bp
- \*       87711    98619: contig of 10909 bp in length
- \*       98620 98719: gap of      100 bp
- \*       98720    108219: contig of 9500 bp in length
- \*       108220 108319: gap of      100 bp
- \*       108320    116259: contig of 7940 bp in length
- \*       116260 116359: gap of      100 bp
- \*       116360    123262: contig of 6903 bp in length
- \*       123263 123362: gap of      100 bp
- \*       123363    128558: contig of 5196 bp in length
- \*       128559 128658: gap of      100 bp
- \*       128659    134632: contig of 5974 bp in length
- \*       134633 134732: gap of      100 bp
- \*       134733    140461: contig of 5729 bp in length
- \*       140462 140561: gap of      100 bp
- \*       140562    146381: contig of 5820 bp in length
- \*       146382 146481: gap of      100 bp
- \*       146482    151457: contig of 4976 bp in length
- \*       151458 151557: gap of      100 bp
- \*       151558    153639: contig of 2082 bp in length
- \*       153640 153739: gap of      100 bp
- \*       153740    157220: contig of 3481 bp in length
- \*       157221 157320: gap of      100 bp
- \*       157321    161091: contig of 3771 bp in length
- \*       161092 161191: gap of      100 bp
- \*       161192    164668: contig of 3477 bp in length
- \*       164669 164768: gap of      100 bp

```

*    164769    168197: contig of 3429 bp in length
*    168198    168297: gap of      100 bp
*    168298    171770: contig of 3473 bp in length
*    171771    171870: gap of      100 bp
*    171871    175319: contig of 3449 bp in length
*    175320    175419: gap of      100 bp
*    175420    178136: contig of 2717 bp in length
*    178137    178236: gap of      100 bp
*    178237    181402: contig of 3166 bp in length
*    181403    181502: gap of      100 bp
*    181503    182931: contig of 1429 bp in length
*    182932    183031: gap of      100 bp
*    183032    184947: contig of 1916 bp in length
*    184948    185047: gap of      100 bp
*    185048    186820: contig of 1773 bp in length
*    186821    186920: gap of      100 bp
*    186921    188144: contig of 1224 bp in length
*    188145    188244: gap of      100 bp
*    188245    189483: contig of 1239 bp in length
*    189484    189583: gap of      100 bp
*    189584    190739: contig of 1156 bp in length.

FEATURES          Location/Qualifiers
source           1. .190739
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="11"
                  /map="11q14"
                  /clone="RP11-853O20"

misc_feature     1. .19670
                  /note="assembly_fragment"
misc_feature     19771. .32416
                  /note="assembly_fragment"
misc_feature     32517. .44331
                  /note="assembly_fragment"
misc_feature     44432. .57404
                  /note="assembly_fragment"
misc_feature     57505. .67680
                  /note="assembly_fragment"
misc_feature     67781. .78637
                  /note="assembly_fragment"
misc_feature     78738. .87610
                  /note="assembly_fragment"
misc_feature     87711. .98619
                  /note="assembly_fragment"
misc_feature     98720. .108219
                  /note="assembly_fragment"
misc_feature     108320. .116259
                  /note="assembly_fragment"
misc_feature     116360. .123262
                  /note="assembly_fragment"
misc_feature     123363..128558
                  /note="assembly_fragment"
misc_feature     128659. .134632
                  /note="assembly_fragment"
misc_feature     134733. .140461
                  /note="assembly_fragment"
misc feature    140562. .146381

```

```

        /note="assembly_fragment"
misc_feature 146482. .151457
        /note="assembly_fragment"
misc_feature 151558. .153639
        /note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature 153740. .157220
        /note="assembly_fragment"
misc_feature 157321. .161091
        /note="assembly_fragment"
misc_feature 161192. .164668
        /note="assembly_fragment"
misc_feature 164769. .168197

Query Match          8.7%;  Score 37.8;  DB 2;  Length 190739;
Best Local Similarity 55.8%;  Pred. No. 7.4;
Matches    72;  Conservative   0;  Mismatches   57;  Indels     0;  Gaps      0;

Qy      68 tggacatctcaaattagcttaaggcgatacatgttatgtccactagagaaaacaacatcc 127
       ||||||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    75761 TGGACATATCTAATTAAAATAACAATATCCATTTGTATATGCAATAAACCTACCTTC 75820

Qy      128 tgagacactcactttatttggaaatgtctcgcgattatcgctgatgtggacatgttta 187
       || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    75821 TGTAAAAGTAACAATTTCTGGTCTGTTGCTGGATAACCACCTAAAATGAGGTTTTT 75880

Qy      188 catgcttct 196
       | | | |
Db    75881 TTTAATTAT 75889

```

RESULT 8  
AP000710  
LOCUS AP000710 108464 bp DNA HTG 30-MAY-2000  
DEFINITION Homo sapiens chromosome 11 clone CMB9-50C9 map 11q25, WORKING DRAFT  
SEQUENCE, 9 unordered pieces.  
ACCESSION AP000710  
VERSION AP000710.2 GI:8118879  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens DNA, clone:CMB9-50C9.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 108464)  
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE Homo sapiens 108,464 genomic DNA of 11q25  
JOURNAL Published Only in DataBase (1999) In press  
REFERENCE 2 (bases 1 to 108464)  
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE Direct Submission  
JOURNAL Submitted (11-NOV-1999) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,  
Japan (E-mail:hattori@gsc.riken.go.jp,  
URL:<http://hgp.gsc.riken.go.jp/>, Tel:81-42-778-9923,  
Fax:81-42-778-9924)

COMMENT On May 31, 2000 this sequence version replaced gi:6997565.

----- Genome Center

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: <http://hgp.gsc.riken.go.jp/>

Contact: hattori@gsc.riken.go.jp

----- Project Information

Center project name: HumDraft11

Center clone name: CMB9-50C9

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 102881 bases at least Q40

Consensus quality: 105715 bases at least Q30

Consensus quality: 106938 bases at least Q20

Insert size: 107664; sum-of-contigs

Quality coverage: 5.58x in Q20 bases; sum-of-contigs

-----.

NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1	47733	contig of	47733 bp	in length
47834	66235	contig of	18402 bp	in length
66336	82251	contig of	15916 bp	in length
82352	90903	contig of	8552 bp	in length
91004	97034	contig of	6031 bp	in length
97135	100846	contig of	3712 bp	in length
100947	104148	contig of	3202 bp	in length
104249	107281	contig of	3033 bp	in length
107382	108464	contig of	1083 bp	in length

Sequence updated (26-May-2000).

\* NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 47733: contig of 47733 bp in length  
\* 47734 47833: gap of 100 bp  
\* 47834 66235: contig of 18402 bp in length  
\* 66236 66335: gap of 100 bp  
\* 66336 82251: contig of 15916 bp in length  
\* 82252 82351: gap of 100 bp  
\* 82352 90903: contig of 8552 bp in length  
\* 90904 91003: gap of 100 bp  
\* 91004 97034: contig of 6031 bp in length  
\* 97035 97134: gap of 100 bp  
\* 97135 100846: contig of 3712 bp in length  
\* 100847 100946: gap of 100 bp  
\* 100947 104148: contig of 3202 bp in length

\* 104149 104248: gap of 100 bp  
 \* 104249 107281: contig of 3033 bp in length  
 \* 107282 107381: gap of 100 bp  
 \* 107382 108464: contig of 1083 bp in length.  
**FEATURES** Location/Qualifiers  
 source 1. .108464  
   /organism="Homo sapiens"  
   /db\_xref="taxon:9606"  
   /chromosome="11"  
   /map="11q25"  
   /clone="CMB9-50C9"  
 misc\_feature 1. .47733  
   /note="assembly\_fragment"  
 misc\_feature 47834. .66235  
   /note="assembly\_fragment"  
 misc\_feature 66336. .82251  
   /note="assembly\_fragment"  
 misc\_feature 82352. .90903  
   /note="assembly\_fragment"  
 misc\_feature 91004. .97034  
   /note="assembly\_fragment"  
 misc\_feature 97135. .100846  
   /note="assembly\_fragment"  
 misc\_feature 100947. .104148  
   /note="assembly\_fragment"  
 misc\_feature 104249. .107281  
   /note="assembly\_fragment"  
 misc\_feature 107382. .108464  
   /note="assembly\_fragment"  
 BASE COUNT 32761 a 19516 c 21313 g 34074 t 800 others  
 ORIGIN  
  
 Query Match 8.6%; Score 37.4; DB 2; Length 108464;  
 Best Local Similarity 52.2%; Pred. No. 8.8;  
 Matches 83; Conservative 0; Mismatches 76; Indels 0; Gaps 0;  
  
 Qy 271 tgtttatattctgctcgacaacgagtatgaaattaaggcacgttatcatcagtgaatgaaac 330  
   | | |||| || | || || | || | || | || | || | || | || | || | || | || |  
 Db 17281 TCTATATTGGTAAGAAAATACCAATTGATTTAGTGAACCTAGCACCAAACCTCAGTA 17340  
  
 Qy 331 agatactaaaatttaatcatttcgctatcgcgattttatatcgtatctgttccatctg 390  
   || | | || | || | || | | || | || | || | || | || | || | || | || |  
 Db 17341 CCATCTGACAGTTCTCTCCTCTGTGGATATCATTAAGTCCTAATAGTTTACCTG 17400  
  
 Qy 391 tcgtgagtgacatcattttattcgccggctttc 429  
   | || | | || | || | || | || | || | || | || | || | || |  
 Db 17401 TGTCATTCGTGCCTTCATTATTCTTACTTCTCTAC 17439

RESULT 9  
 HS297A17/c  
 LOCUS HS297A17 160759 bp DNA PRI 20-JUL-2001  
 DEFINITION Homo sapiens chromosome 9 BAC RP11-297A17, complete sequence.  
 ACCESSION AL513503 AL353134  
 VERSION AL513503.1 GI:12733884  
 KEYWORDS HTG.

SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 160759)  
AUTHORS Jaerke,D., Conrad,A., Hornischer,K., Loehnert,T.H., Scharfe,M.,  
Thies,S. and Bloecker,H.  
TITLE Direct Submission  
JOURNAL Submitted (07-FEB-2001) GBF, Dept. of Genome Analysis, Mascheroder  
Weg 1, D-38124 Braunschweig, Germany, E-mail: info.genome@gbf.de  
COMMENT On Mar 21, 2001 this sequence version replaced gi:12330759.  
All annotations in this database entry are developed by  
computational tools. It is therefore not explicitly noted in the  
feature lines that evidence is not experimental.  
Mapping was performed at The Sanger Centre  
(cf. <http://www.sanger.ac.uk/HGP/Chr9>)  
Mapping information is available via  
<http://webace.sanger.ac.uk/cgi-bin/display?db=acedb9&grep=297A17>  
----- Genome Center  
Center: GBF, Braunschweig  
Center code: GBF  
Web site: <http://genome.gbf.de>  
Contact: info.genome@gbf.de  
----- Project Information  
Center project name:  
Center clone name: 297A17  
----- Summary Statistics  
Sequencing vector: ###;  
Chemistry: Dye-terminator-BigDye: 77,9% of reads  
Chemistry: Dye-terminator-amersham: 20,7% of reads Chemistry:  
Dye-primer-amersham: 1,4% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 0 bases at least Q40  
Consensus quality: 0 bases at least Q30  
Consensus quality: 0 bases at least Q20  
Estimated insert size: ##; agarose-fp estimation  
Estimated insert size: 160759; sum-of-contigs estimation  
-----  
PROGRAMS AND PARAMETERS USED FOR ANNOTATION:  
+++++  
+ Analysis and annotation were performed with the automatic +  
+ 'first-pass' annotation and submission tool +  
+ 'AnnoMitter' (Hornischer & Bloecker). +  
+ Programs used by 'AnnoMitter': +  
+++++  
> GeneFinder (Green), Vers. 084  
. Organism: human  
> GenScan (Burge & Karlin), Vers. 1.0  
. Used matrix: vertebrate; Minimum score: 0  
> Grail (Xu et al.), Vers. 1.3  
. Organism: human  
> Mzef (Zhang)  
. Prior probability: 0.04; Overlapping number: 0 > Xpound (Thomas  
& Skolnick)  
. Base score cutoff: 0.2; Minimal exon length: 3 bp > 'Repeats':  
BLASTN 2.0.14 (Altschul et al.)  
. Database(s): \* RepBase: ALU (human), released 22-DEC-1995 .

```

        * RepBase: THR ((human), released 22-DEC-1995 .
        * RepBase: L1 (primate), released 22-DEC-1995 .
        * RepBase: MIR (primate), released 22-DEC-1995 .
        * RepBase: MER (primate), released 22-DEC-1995 .
        * RepBase: MIR2 (primate), released 22-DEC-1995 .
        * RepBase: THE (primate), released 22-DEC-1995 . Minimum score: 60;
        Minimum identity: 70 %;
        > 'ESTs': BLASTN 2.0.14 (Altschul et al.)
        . Database(s): * embl (EST, human), released -DEC- .
        . * embl (EST, other), released -DEC- . * emblnew
        (EST), released -DEC-
        . Using sequence with masked repeats
        . Minimum score: 60; Minimum identity: 70 %;
        > 'Tandem Repeats': GDE 2.2 option 'tandem'
        . Minimum length 2 bp; Maximum length 20 bp; Score threshold 20 .
        Treat N's as mismatches? YES; Allow uniform consensi? NO >
        'Inverted Repeats': GDE 2.2 option 'inverted'
        > 'Micro Satellites': GDE 2.2 option 'sputnik' (Abajian) > 'CpG
        Islands': GDE 2.2 option 'cpg'
        . CpG island region size 100 bp;
        . Minimum GC contents 50 %; Observed/Expected 0.6 > 'STS Scan':
        e-PCR (Schuler)
        . Margin: 50; Number of mismatches allowed: 0; Word size: 7 .
        STS database: 'dbSTS markers'
        > 'tRNA Scan': tRNAscan-SE (Lowe & Eddy), Vers. 1.11.

FEATURES          Location/Qualifiers
source            1. .160759
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="9"
                  /clone="RP11-297A17"
misc_feature      1. .160759
                  /note="assembly_fragment~clone_end:T7~vector_side:left
                  assembly_fragment~clone_end:SP6~vector_side:right"
exon              521. .570
                  /note="MZEF prediction, score = 0.887"
repeat_region     589. .632
                  /note="homology = 77.3%, counts = 11"
                  /rpt_family="ttcc repeat"
                  /rpt_type=TANDEM
exon              complement(1038. .1136)
                  /note="GENSCAN prediction, score = 2.21
                  MZEF prediction, score = 0.553"
repeat_region     1470. .1531
                  /note="homology = 80.6%, counts = 31"
                  /rpt_family="gt repeat"
                  /rpt_type=TANDEM
satellite         1503. .1531
                  /note="TG repeat"
exon              2107. .2273
                  /note="MZEF prediction, score = 0.661"
repeat_region     complement(2813. .2846)
                  /note="97% identity: matches 22. .55 of consensus"
                  /rpt_family="THE"
exon              complement(4001. .4132)
                  /note="XPOUND prediction, score = 0.680"
exon              complement(4885. .5013)

```

```
        /note="GRAIL, score = 61%, comment = good"
repeat_region
complement(6079. .6207)
/note="95% identity: matches 1166. .1294 of consensus"
/rpt_family="L1"
exon
6143. .6148
/note="XPOUND prediction, score = 0.240"
repeat_region
complement(6224. .7808)
/note="95% identity: matches 401. .1984 of consensus"
/rpt_family="L1"
exon
complement(6396. .7142)
/note="GRAIL, score = 52%, comment = good"
exon
complement(6724. .7594)
/note="GENSCAN prediction, score = 2.26"
exon
6808. .6820
/note="XPOUND prediction, score = 0.215"
complement(7464. .7785)
/note="GRAIL, score = 50%, comment = good"
repeat_region
7835. .8112
/note="91% identity: matches 248. .526 of consensus"
/rpt_family="L1"
repeat_region
complement(7835. .8112)
/note="98% identity: matches 1. .278 of consensus"
/rpt_family="ALU"
misc_feature
7958. .8109
/note="CpG_island (%GC=57.9, o/e=1.08, #CpGs=13)"
repeat_region
complement(8113. .10347)
/note="95% identity: matches 5. .2243 of consensus"
/rpt_family="L1"
exon
complement(8734. .9085)
/note="GRAIL, score = 40%, comment = marginal"
exon
complement(8734. .9175)
/note="GENSCAN prediction, score = 28.27"
exon
complement(9250. .9378)
/note="GRAIL, score = 68%, comment = good"
exon
9563. .9588
/note="XPOUND prediction, score = 0.376"
exon
9952. .9989
/note="XPOUND prediction, score = 0.260"
exon
complement(10747. .10764)
/note="XPOUND prediction, score = 0.393"
exon
complement(10823. .10836)
/note="XPOUND prediction, score = 0.213"
satellite
11136. .11148
/note="CATT repeat"
exon
complement(11380. .11558)
/note="GRAIL, score = 68%, comment = good"
exon
complement(11384. .11558)
/note="MZEf prediction, score = 0.840"
repeat_region
11921. .11980
/note="homology = 73.3%, counts = 20"
/rpt_family="tat repeat"
/rpt_type=TANDEM
satellite
11931. .11942
/note="ATT repeat"
exon
complement(12286. .12329)
/note="MZEf prediction, score = 0.763"
STS
12369. .12510
```

```
/standard_name="SHGC-16985 (D1S1563), Map: 9, Homo  
sapiens"  
/note="GenBank Accession Number: G15514"  
12398. .12615  
/standard_name="TIGR-A003M39 (D12S1978), Map: 62.1, Homo  
sapiens"  
/note="GenBank Accession Number: G26344"  
12996. .13050  
/note="MZEf prediction, score = 0.941"  
STS 14069. .14261  
/standard_name="A005033 (D12S8), Map: 6p11, Homo sapiens"  
/note="GenBank Accession Number: G20382"  
exon complement(15641. .15678)  
/note="MZEf prediction, score = 0.545"  
exon complement(15867. .15937)  
/note="MZEf prediction, score = 0.677"  
repeat_region complement(16090. .16214)  
/note="92% identity: matches 5382. .5506 of consensus"  
repeat_region /rpt_family="L1"  
complement(16254. .16370)  
/note="91% identity: matches 314. .430 of consensus"  
repeat_region /rpt_family="L1"  
16369. .16801  
/note="94% identity: matches 325. .758 of consensus"
```

Query Match 8.6%; Score 37.4; DB 9; Length 160759;  
Best Local Similarity 52.2%; Pred. No. 9.3;  
Matches 83; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

```

Qy      331 agatactaaaatttaatcatttcgctatcgcgattttatatatcgatctgttccatctg 390
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Pb 153751 CCATCTGACAGTTTCTCTCCTTCTGTGGATATCATTAAAGTCCTAATAGTTTACCTG 153692

```

```

Qy      391 tcgtgagtgtgacatcattttattcgccggcttc 429
          |       |||   | | | | | | | | | | | | | |
Db 153691 TGTCAATTCTGTCGCCTCATTTTATTCTTACTTCTCTAC 153653

```

RESULT 10  
HS520K3  
LOCUS HS520K3 176210 bp DNA HTG 07-MAY-2001  
DEFINITION Homo sapiens chromosome 9 clone RP11-520K3, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 8 unordered pieces.  
ACCESSION AL450004 AL162251  
VERSION AL450004.1 GI:11138112  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_ACTIVEFIN; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 176210)  
AUTHORS Plumb, B.  
TITLE Direct Submission

JOURNAL Submitted (19-APR-2000) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone  
requests: clonerequest@sanger.ac.uk

REFERENCE 2 (bases 1 to 176210)

AUTHORS Nordsiek,G., Conrad,A., Hornischer,K., Loehnert,T.H., Scharfe,M.,  
Schoen,O. and Bloecker,H.

TITLE Direct Submission

JOURNAL Submitted (09-NOV-2000) GBF, Dept. of Genome Analysis, Mascheroder  
Weg 1, D-38124 Braunschweig, Germany, E-mail: info.genome@gbf.de

COMMENT On May 14, 2001 this sequence version replaced gi:9212909.  
All annotations in this database entry are developed by  
computational tools. It is therefore not explicitly noted in the  
feature lines that evidence is not experimental.  
Mapping was performed at The Sanger Centre  
(cf. <http://www.sanger.ac.uk/HGP/Chr9>)  
Mapping information is available via  
<http://webace.sanger.ac.uk/cgi-bin/display?db=acedb9&grep=520K3>  
----- Genome Center  
Center: GBF, Braunschweig  
Center code: GBF  
Web site: <http://genome.gbf.de/>  
Contact: info.genome@gbf.de  
----- Project Information  
Center project name:  
Center clone name: bA520K3  
----- Summary Statistics  
Sequencing vector: pUC18;  
Chemistry: Dye-terminator-BigDye: 58% of reads  
Chemistry: Dye-terminator-ámersham: 42% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 145030 bases at least Q40  
Consensus quality: 146258 bases at least Q30  
Consensus quality: 146930 bases at least Q20  
Estimated insert size: 175510; sum-of-contigs estimation  
-----  
PROGRAMS AND PARAMETERS USED FOR ANNOTATION:  
+++++  
+ Analysis and annotation were performed with the automatic +  
+ 'first-pass' annotation and submission tool +  
+ 'AnnoMitter' (Hornischer & Bloecker). +  
+ Programs used by 'AnnoMitter': +  
+++++  
> GeneFinder (Green), Vers. 084  
. Organism: human  
> GenScan (Burge & Karlin), Vers. 1.0  
. Used matrix: vertebrate; Minimum score: 0  
> Grail (Xu et al.), Vers. 1.3  
. Organism: human  
> Mzef (Zhang)  
. Prior probability: 0.04; Overlapping number: 0 > Xpound (Thomas  
& Skolnick)  
. Base score cutoff: 0.2; Minimal exon length: 3 bp > 'Repeats':  
BLASTN 2.0.14 (Altschul et al.)  
. Database(s): \* RepBase: ALU (human), released 22-DEC-1995 .  
                  \* RepBase: THR ((human), released 22-DEC-1995 .  
                  \* RepBase: L1 (primate), released 22-DEC-1995 .  
\* RepBase: MIR (primate), released 22-DEC-1995 . \*

```

RepBase: MER (primate), released 22-DEC-1995 . *
RepBase: MIR2 (primate), released 22-DEC-1995 . *
RepBase: THE (primate), released 22-DEC-1995 . Minimum score: 60;
Minimum identity: 70 %;
> 'Tandem Repeats': GDE 2.2 option 'tandem'
. Minimum length 2 bp; Maximum length 20 bp; Score threshold 20 .
Treat N's as mismatches? YES; Allow uniform consensi? NO >
'Inverted Repeats': GDE 2.2 option 'inverted'
> 'Micro Satellites': GDE 2.2 option 'sputnik' (Abajian) > 'CpG
Islands': GDE 2.2 option 'cpg'
. Cpg island region size 100 bp;
. Minimum GC contents 50 %; Observed/Expected 0.6 > 'STS Scan':
e-PCR (Schuler)
. Margin: 50; Number of mismatches allowed: 0; Word size: 7 .
STS database: 'dbSTS markers'
> 'tRNA Scan': tRNAscan-SE (Lowe & Eddy), Vers. 1.11.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*      1    28146: contig of 28146 bp in length
*      28147 28246: gap of      100 bp
*      28247    74684: contig of 46438 bp in length
*      74685 74784: gap of      100 bp
*      74785    86091: contig of 11307 bp in length
*      86092 86191: gap of      100 bp
*      86192    88464: contig of 2273 bp in length
*      88465 88564: gap of      100 bp
*      88565    93628: contig of 5064 bp in length
*      93629 93728: gap of      100 bp
*      93729    151082: contig of 57354 bp in length
*      151083 151182: gap of      100 bp
*      151183    162592: contig of 11410 bp in length
*      162593 162692: gap of      100 bp
*      162693    176210: contig of 13518 bp in length.

FEATURES
source          Location/Qualifiers
source          1. .176210
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="9"
                /clone="RP11-520K3"
                /clone_lib="RPCI-11.2"
misc_feature    1. .28146
                /note="assembly_fragment
clone_end:T7
vector_side:left"
exon            303. .375
                /note="MZF prediction, score = 0.709"
exon            1120. .1172
                /note="MZF prediction, score = 0.937"
repeat_region   1210. .1241
                /note="homology = 93.8%, counts = 4"
                /rpt_family="aattgaat repeat"

```

```
/rpt_type=TANDEM
exon 1299. .1405
/repeat_region /note="MZE prediction, score = 0.777"
complement(1315. .1368)
/note="90% identity: matches 199. .252 of consensus"
/rpt_family="L1"
satellite 1996. .2009
/exon /note="CA repeat"
complement(2434. .2536)
/note="GRAIL, score = 98%, comment = excellent
MZE prediction, score = 0.923"
complement(3301. .3465)
/exon /note="GRAIL, score = 52%, comment = good"
complement(4201. .4262)
/note="GRAIL, score = 68%, comment = good"
complement(4201. .4251)
/exon /note="XPOUND prediction, score = 0.390"
complement(4343. .4429)
/exon /note="MZE prediction, score = 0.820"
complement(4525. .4661)
/repeat_region /note="IR1, 77% complementary to IR1' (5609. .5745)"
/rpt_type=INVERTED
repeat_region 4536. .4803
/note="81% identity: matches 161. .424 of consensus"
/rpt_family="L1"
repeat_region complement(4536. .4808)
/note="84% identity: matches 9. .278 of consensus"
/rpt_family="ALU"
exon complement(4709. .4844)
/note="MZE prediction, score = 0.766"
exon complement(4719. .4750)
/note="XPOUND prediction, score = 0.355"
exon complement(4807. .4816)
/note="XPOUND prediction, score = 0.238"
repeat_region 5609. .5745
/note="IR1', 77% complementary to IR1 (4525. .4661)
83% identity: matches 157. .293 of consensus"
/rpt_family="ALU"
/rpt_type=INVERTED
repeat_region complement(5609. .5727)
/note="81% identity: matches 168. .286 of consensus"
/rpt_family="L1"
exon 6079. .6198
/note="MZE prediction, score = 0.512"
repeat_region 6108. .6270
/note="90% identity: matches 1. .163 of consensus"
/rpt_family="ALU"
repeat_region complement(6108. .6270)
/note="89% identity: matches 363. .526 of consensus"
/rpt_family="L1"
repeat_region 6270. .6379
/note="92% identity: matches 1258. .1367 of consensus"
/rpt_family="ALU"
repeat_region complement(6270. .6379)
/note="88% identity: matches 243. .352 of consensus"
/rpt_family="L1"
repeat_region 6378. .6529
```

```

repeat_region          /note="homology = 64.5%, counts = 8"
repeat_region          /rpt_family="aataaagaaagaaaaaaaa repeat"
repeat_region          /rpt_type=TANDEM
repeat_region          6674. .6689
repeat_region          /note="IR2, 100% complementary to IR2' (6692. .6707)"
repeat_region          /rpt_type=INVERTED
repeat_region          6677. .6704
repeat_region          /note="homology = 100.0%, counts = 14"
repeat_region          /rpt_family="at repeat"
repeat_region          /rpt_type=TANDEM
repeat_region          6692. .6707
repeat_region          /note="IR2', 100% complementary to IR2 (6674. .6689)"
repeat_region          /rpt_type=INVERTED
exon                 8640. .8747
exon                 /note="GRAIL, score = 41%, comment = marginal"
exon                 9417. .9636

Query Match           8.6%; Score 37.4; DB 2; Length 176210;
Best Local Similarity 52.2%; Pred. No. 9.4;
Matches   83; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy      271 tgtttatattctgctcgacaacgagttatcgttatacgtgaatgaaac 330
        ||||| || | || ||| | ||| | ||| | | | | | | | | | | | | |
Db      54762 TCTATATTGGTAAGAAAATCACCATTGACTTACCAACCTCAGTA 54821

Qy      331 agataactaaaatttaatcatttcgctatcgcgattttatatcgtatctgtccatcg 390
        || | | | || | || | | | | | | | | | | | | | | | | | | |
Db      54822 CCATCTGACAGTTCTCTCCTCTGTGGATATCATTAAGTCCTAATAGTTTACCTG 54881

Qy      391 tcgtgagttgtgacatcattttattcgtccggcttc 429
        | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      54882 TGTCATTCGTGCCTTCATTATTCTTACTTCTCTAC 54920

```

```

RESULT 11
CER01H5
LOCUS    CER01H5      22887 bp      DNA          INV      25-OCT-2000
DEFINITION Caenorhabditis elegans cosmid R01H5, complete sequence.
ACCESSION Z68007
VERSION   Z68007.1  GI:1070077
KEYWORDS  HTG; Thr-tRNA; Transfer RNA.
SOURCE    Caenorhabditis elegans.
ORGANISM  Caenorhabditis elegans
           Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
           Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 22887)
AUTHORS   none.
TITLE     Genome sequence of the nematode C. elegans: a platform for
           investigating biology. The C. elegans Sequencing Consortium
JOURNAL   Science 282 (5396), 2012-2018 (1998)
MEDLINE   99069613
REMARK    The C.elegans Sequencing Consortium.
           Erratum:[ [published errata appear in Science 1999 Jan
           1;283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep
           3;285(5433):1493] ]
REFERENCE 2 (bases 1 to 22887)
AUTHORS   Lloyd,C.R.

```

**TITLE** Direct Submission  
**JOURNAL** Submitted (21-NOV-1995) Nematode Sequencing Project, Sanger Centre,  
 Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,  
 Washington University, St. Louis, MO 63110, USA. E-mail:  
 jes@sanger.ac.uk or rw@nematode.wustl.edu  
**COMMENT** Coding sequences below are predicted from computer analysis, using  
 predictions from Genefinder (P. Green, U. Washington), and other  
 available information.  
 Current sequence finishing criteria for the *C. elegans* genome  
 sequencing consortium are that all bases are either sequenced  
 unambiguously on both strands, or on a single strand with both a  
 dye primer and dye terminator reaction, from distinct subclones.  
 Exceptions are indicated by an explicit note.  
**IMPORTANT:** This sequence is not the entire insert of clone R01H5.  
 It may be shorter because we only sequence overlapping sections  
 once, or longer because we arrange for a small overlap between  
 neighbouring submissions.  
 The true left end of clone R01H5 is at 1 in this sequence. The true  
 right end of clone R01H5 is at 18898 in  
 sequence Z68012.  
 The true left end of clone T24D5 is at 22788 in this sequence. The  
 true right end of clone C03C5 is at 6810 in this sequence. The  
 start of this sequence (1..104) overlaps with the end of sequence  
 Z81472.  
 The end of this sequence (22788..22887) overlaps with the start of  
 sequence Z68012.  
 For a graphical representation of this sequence and its analysis  
 see:- [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?  
 name=R01H5](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=R01H5).

**FEATURES**  
**source** Location/Qualifiers  
 1..22887  
 /organism="Caenorhabditis elegans"  
 /db\_xref="taxon:6239"  
 /chromosome="X"  
 /clone="R01H5"  
**tRNA** 20300..20371  
 /gene="R01H5.t1"  
 /note="TGT Thr T-tRNA  
 predicted using tRNAscan-SE-1.11  
 preliminary prediction  
 similar to tRNA-Thr"  
**gene** 20300..20371  
 /gene="R01H5.t1"

<b>BASE COUNT</b>	8101	a	3574	c	3994	g	7218	t
<b>ORIGIN</b>								

Query Match 8.5%; Score 36.8; DB 3; Length 22887;  
Best Local Similarity 46.0%; Pred. No. 11;  
Matches 125; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

Qy 204 taaaagtctttgctccgaatctcgagacgagatttttaaaaaaaaaaaaaaa 263

```

Db 15761 CTGAAACTGAATTTAAAAAATCTTTATTCCGGTCAATATCTGAAATTGACATTCAA 15820
Qy 264 ccccagggtgttatattctgctcgacaacgagttatggaaattaaggcacgttatcatcgta 323
     | | | | | || | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15821 AAGAAAATAAAAACATGCTACGGAAATGCTAATTATGACATCAGCTATTGTTAGCATTT 15880
Qy 324 atgaaaacagatactaaaattaatcatttcgctatcgcgattttatatcgatctgtt 383
     | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15881 TTAATACATAATATTTATTAGAAATTGACATTTCAGAGTTCTAAACTATGTGTT 15940
Qy 384 ccatctgtcgtagtgacatcattttatt 415
     | | | | | | | | | | | | | | | | | |
Db 15941 TCAACAAACGTGAACATGTCATTTCCTTTT 15972

```

RESULT 12  
AC084841/c  
LOCUS AC084841 72356 bp DNA HTG 09-MAY-2001  
DEFINITION Homo sapiens chromosome 8 clone CTD-2131E13 map 8, WORKING DRAFT  
SEQUENCE, 2 ordered pieces.  
ACCESSION AC084841  
VERSION AC084841.2 GI:13940656  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 72356)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens chromosome 8, clone CTD-2131E13  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 72356)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,  
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,  
FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,  
Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,  
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,  
Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,  
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,  
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,  
O'Donnell,P., O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K.,  
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,  
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,  
Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,  
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,  
Zimmer,A. and Zody,M.  
TITLE Direct Submission  
JOURNAL Submitted (22-NOV-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On May 4, 2001 this sequence version replaced gi:11276198.  
All repeats were identified using RepeatMasker.



```

Qy      230 gacgagatttttaaggggggagggtctgtacaccccaagggtgttatattctgctcgac 289
       | | | ||||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db      48013 TAATACTATATTTATACTTGGAGTTTATAAGACAGTAGATCTAAATGTTTCACCAC 47954

Qy      290 aacgagtatggaattaagcacgttatcatcagtgaatgaaacagatact. 337
       | | | . | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      47953 ACACACAAAAATCATATCGTGTGAGGTGAGAGATTGTGGCAATTATT 47906

```

RESULT 13  
AC009197  
LOCUS AC009197 114084 bp DNA HTG 31-JAN-2000  
DEFINITION Drosophila melanogaster chromosome 2 clone BACR14M08 (D1019)  
RPCI-98 14.M.8 map 30A-30E strain y; cn bw sp, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 80 unordered pieces.  
ACCESSION AC009197  
VERSION AC009197.7 GI:6838840  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydriodea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 114084)  
AUTHORS Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,  
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,  
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,  
Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,  
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.  
TITLE Sequencing of Drosophila melanogaster  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 114084)  
AUTHORS Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,  
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,  
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,  
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.  
TITLE Direct Submission  
JOURNAL Submitted (06-AUG-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA  
COMMENT On Jan 31, 2000 this sequence version replaced gi:6806805.  
For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [bdgp@fruitfly.berkeley.edu](mailto:bdgp@fruitfly.berkeley.edu). All contigs in this submission meet the following cutoffs: length >= 200 bases.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 80 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 757: contig of 757 bp in length  
\* 758 837: gap of unknown length  
\* 838 1748: contig of 911 bp in length  
\* 1749 1828: gap of unknown length  
\* 1829 2477: contig of 649 bp in length  
\* 2478 2557: gap of unknown length  
\* 2558 3493: contig of 936 bp in length  
\* 3494 3573: gap of unknown length  
\* 3574 4230: contig of 657 bp in length  
\* 4231 4310: gap of unknown length  
\* 4311 4952: contig of 642 bp in length  
\* 4953 5032: gap of unknown length  
\* 5033 6086: contig of 1054 bp in length  
\* 6087 6166: gap of unknown length  
\* 6167 7490: contig of 1324 bp in length  
\* 7491 7570: gap of unknown length  
\* 7571 8162: contig of 592 bp in length  
\* 8163 8242: gap of unknown length  
\* 8243 8909: contig of 667 bp in length  
\* 8910 8989: gap of unknown length  
\* 8990 10012: contig of 1023 bp in length  
\* 10013 10092: gap of unknown length  
\* 10093 11156: contig of 1064 bp in length  
\* 11157 11236: gap of unknown length  
\* 11237 12057: contig of 821 bp in length  
\* 12058 12137: gap of unknown length  
\* 12138 13660: contig of 1523 bp in length  
\* 13661 13740: gap of unknown length  
\* 13741 14935: contig of 1195 bp in length  
\* 14936 15015: gap of unknown length  
\* 15016 15923: contig of 908 bp in length  
\* 15924 16003: gap of unknown length  
\* 16004 17276: contig of 1273 bp in length  
\* 17277 17356: gap of unknown length  
\* 17357 18126: contig of 770 bp in length  
\* 18127 18206: gap of unknown length  
\* 18207 19111: contig of 905 bp in length  
\* 19112 19191: gap of unknown length  
\* 19192 20232: contig of 1041 bp in length  
\* 20233 20312: gap of unknown length  
\* 20313 21209: contig of 897 bp in length  
\* 21210 21289: gap of unknown length  
\* 21290 22416: contig of 1127 bp in length  
\* 22417 22496: gap of unknown length  
\* 22497 23541: contig of 1045 bp in length  
\* 23542 23621: gap of unknown length  
\* 23622 24876: contig of 1255 bp in length  
\* 24877 24956: gap of unknown length  
\* 24957 25789: contig of 833 bp in length  
\* 25790 25869: gap of unknown length  
\* 25870 27270: contig of 1401 bp in length  
\* 27271 27350: gap of unknown length

\* 27351 28905: contig of 1555 bp in length  
\* 28906 28985: gap of unknown length  
\* 28986 30878: contig of 1893 bp in length  
\* 30879 30958: gap of unknown length  
\* 30959 32106: contig of 1148 bp in length  
\* 32107 32186: gap of unknown length  
\* 32187 33070: contig of 884 bp in length  
\* 33071 33150: gap of unknown length  
\* 33151 34933: contig of 1783 bp in length  
\* 34934 35013: gap of unknown length  
\* 35014 37059: contig of 2046 bp in length  
\* 37060 37139: gap of unknown length  
\* 37140 38293: contig of 1154 bp in length  
\* 38294 38373: gap of unknown length  
\* 38374 40181: contig of 1808 bp in length  
\* 40182 40261: gap of unknown length  
\* 40262 41039: contig of 778 bp in length  
\* 41040 41119: gap of unknown length  
\* 41120 43453: contig of 2334 bp in length  
\* 43454 43533: gap of unknown length  
\* 43534 45410: contig of 1877 bp in length  
\* 45411 45490: gap of unknown length  
\* 45491 47459: contig of 1969 bp in length  
\* 47460 47539: gap of unknown length  
\* 47540 49956: contig of 2417 bp in length  
\* 49957 50036: gap of unknown length  
\* 50037 52295: contig of 2259 bp in length  
\* 52296 52375: gap of unknown length  
\* 52376 55196: contig of 2821 bp in length  
\* 55197 55276: gap of unknown length  
\* 55277 58005: contig of 2729 bp in length  
\* 58006 58085: gap of unknown length  
\* 58086 60342: contig of 2257 bp in length  
\* 60343 60422: gap of unknown length  
\* 60423 65739: contig of 5317 bp in length  
\* 65740 65819: gap of unknown length  
\* 65820 70797: contig of 4978 bp in length  
\* 70798 70877: gap of unknown length  
\* 70878 77297: contig of 6420 bp in length  
\* 77298 77377: gap of unknown length  
\* 77378 84770: contig of 7393 bp in length  
\* 84771 84850: gap of unknown length  
\* 84851 90631: contig of 5781 bp in length  
\* 90632 90711: gap of unknown length  
\* 90712 91301: contig of 590 bp in length  
\* 91302 91381: gap of unknown length  
\* 91382 92015: contig of 634 bp in length  
\* 92016 92095: gap of unknown length  
\* 92096 92756: contig of 661 bp in length  
\* 92757 92836: gap of unknown length  
\* 92837 93488: contig of 652 bp in length  
\* 93489 93568: gap of unknown length  
\* 93569 94252: contig of 684 bp in length  
\* 94253 94332: gap of unknown length  
\* 94333 95046: contig of 714 bp in length  
\* 95047 95126: gap of unknown length  
\* 95127 95824: contig of 698 bp in length

\* 95825 95904: gap of unknown length  
\* 95905 96643: contig of 739 bp in length  
\* 96644 96723: gap of unknown length  
\* 96724 97433: contig of 710 bp in length  
\* 97434 97513: gap of unknown length  
\* 97514 98221: contig of 708 bp in length  
\* 98222 98301: gap of unknown length  
\* 98302 98905: contig of 604 bp in length  
\* 98906 98985: gap of unknown length  
\* 98986 99440: contig of 455 bp in length  
\* 99441 99520: gap of unknown length  
\* 99521 100182: contig of 662 bp in length  
\* 100183 100262: gap of unknown length  
\* 100263 101034: contig of 772 bp in length  
\* 101035 101114: gap of unknown length  
\* 101115 101741: contig of 627 bp in length  
\* 101742 101821: gap of unknown length  
\* 101822 102440: contig of 619 bp in length  
\* 102441 102520: gap of unknown length  
\* 102521 103154: contig of 634 bp in length  
\* 103155 103234: gap of unknown length  
\* 103235 103978: contig of 744 bp in length  
\* 103979 104058: gap of unknown length  
\* 104059 104620: contig of 562 bp in length  
\* 104621 104700: gap of unknown length  
\* 104701 105430: contig of 730 bp in length  
\* 105431 105510: gap of unknown length  
\* 105511 106004: contig of 494 bp in length  
\* 106005 106084: gap of unknown length  
\* 106085 106627: contig of 543 bp in length  
\* 106628 106707: gap of unknown length  
\* 106708 107377: contig of 670 bp in length  
\* 107378 107457: gap of unknown length  
\* 107458 108046: contig of 589 bp in length  
\* 108047 108126: gap of unknown length  
\* 108127 108668: contig of 542 bp in length  
\* 108669 108748: gap of unknown length  
\* 108749 109445: contig of 697 bp in length  
\* 109446 109525: gap of unknown length  
\* 109526 110014: contig of 489 bp in length  
\* 110015 110094: gap of unknown length

Query Match 8.4%; Score 36.6; DB 2; Length 114084;  
Best Local Similarity 34.8%; Pred. No. 15;  
Matches 93; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

Qy 88 taaggcgatacatgttatgtccactagagaacaacatcctgagacactcaccttattt 147  
||||| ||||| ||||| | ||||| | ||||| | | | | | | | | | | | | | | | | | |

```

Qy      148 ggaatgtctcgcatatcgctgatgtggacatgtgttacatgcttctactctaaa 207
       | ||| | ||| | |||| | | | | | | | | | | | | | | | | | | | | | |
Db      47394 AAATATTTTCTCTTTAACAAATAGTTCCCATTCTGCAACTGATTCTACTCAACTATA 47453

```

RESULT 14  
AC025359/c  
LOCUS AC025359 134580 bp DNA HTG 27-APR-2000  
DEFINITION Homo sapiens chromosome 13 clone RP11-354D13 map 13, WORKING DRAFT  
SEQUENCE, 22 unordered pieces.  
ACCESSION AC025359  
VERSION AC025359.3 GI:7656790  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 134580)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens chromosome 13, clone RP11-354D13  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 134580)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
Boguslavkiy,L., Boukhalter,B., Brown,A., Burkett,G.,  
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,  
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
Vassilieff,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Apr 27, 2000 this sequence version replaced gi:7342149.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7802  
Center clone name: 354\_D\_13

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 121112 bases at least Q40  
Consensus quality: 127983 bases at least Q30  
Consensus quality: 130794 bases at least Q20  
Insert size: 147000; agarose-fp  
Insert size: 132480; sum-of-contigs  
Quality coverage: 3.4 in Q20 bases; agarose-fp  
Quality coverage: 3.8 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 22 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1069: contig of 1069 bp in length  
\* 1070 1169: gap of 100 bp  
\* 1170 2333: contig of 1164 bp in length  
\* 2334 2433: gap of 100 bp  
\* 2434 3924: contig of 1491 bp in length  
\* 3925 4024: gap of 100 bp  
\* 4025 5207: contig of 1183 bp in length  
\* 5208 5307: gap of 100 bp  
\* 5308 7841: contig of 2534 bp in length  
\* 7842 7941: gap of 100 bp  
\* 7942 9438: contig of 1497 bp in length  
\* 9439 9538: gap of 100 bp  
\* 9539 13191: contig of 3653 bp in length  
\* 13192 13291: gap of 100 bp  
\* 13292 18133: contig of 4842 bp in length  
\* 18134 18233: gap of 100 bp  
\* 18234 22320: contig of 4087 bp in length  
\* 22321 22420: gap of 100 bp  
\* 22421 27367: contig of 4947 bp in length  
\* 27368 27467: gap of 100 bp  
\* 27468 30653: contig of 3186 bp in length  
\* 30654 30753: gap of 100 bp  
\* 30754 34208: contig of 3455 bp in length  
\* 34209 34308: gap of 100 bp  
\* 34309 41350: contig of 7042 bp in length  
\* 41351 41450: gap of 100 bp  
\* 41451 46299: contig of 4849 bp in length  
\* 46300 46399: gap of 100 bp  
\* 46400 52614: contig of 6215 bp in length  
\* 52615 52714: gap of 100 bp  
\* 52715 59921: contig of 7207 bp in length

\* 59922 60021: gap of 100 bp  
\* 60022 65019: contig of 4998 bp in length  
\* 65020 65119: gap of 100 bp  
\* 65120 74956: contig of 9837 bp in length  
\* 74957 75056: gap of 100 bp  
\* 75057 86958: contig of 11902 bp in length  
\* 86959 87058: gap of 100 bp  
\* 87059 97235: contig of 10177 bp in length  
\* 97236 97335: gap of 100 bp  
\* 97336 112962: contig of 15627 bp in length  
\* 112963 113062: gap of 100 bp  
\* 113063 134580: contig of 21518 bp in length.

FEATURES Location/Qualifiers

source 1. .134580  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/map="13"  
/clone="RP11-354D13"  
/clone\_lib="RPCI-11 Human Male BAC"

misc\_feature 1. .1069  
/note="assembly\_fragment"

misc\_feature 1170. .2333  
/note="assembly\_fragment"

misc\_feature 2434. .3924  
/note="assembly\_fragment"

misc\_feature 4025. .5207  
/note="assembly\_fragment"

misc\_feature 5308. .7841  
/note="assembly\_fragment"

misc\_feature 7942. .9438  
/note="assembly\_fragment"

misc\_feature 9539. .13191  
/note="assembly\_fragment"

misc\_feature 13292. .18133  
/note="assembly\_fragment"

misc\_feature 18234. .22320  
/note="assembly\_fragment"

misc\_feature 22421. .27367  
/note="assembly\_fragment"

misc\_feature 27468. .30653  
/note="assembly\_fragment"

misc\_feature 30754. .34208  
/note="assembly\_fragment"

misc\_feature 34309. .41350  
/note="assembly\_fragment"

clone\_end:T7  
vector\_side:left"

misc\_feature 41451. .46299  
/note="assembly\_fragment"

misc\_feature 46400. .52614  
/note="assembly\_fragment"

misc\_feature 52715. .59921  
/note="assembly\_fragment"

misc\_feature 60022. .65019  
/note="assembly\_fragment"

misc\_feature 65120. .74956

```

        /note="assembly_fragment"
misc_feature    75057. .86958
        /note="assembly_fragment"
misc_feature    87059. .97235
        /note="assembly_fragment"
misc_feature    97336. .112962
        /note="assembly_fragment"
misc_feature    113063. .134580
        /note="assembly_fragment"
clone_end:SP6
vector_side:right"
BASE COUNT      42123 a  24933 c  22813 g  42609 t   2102 others
ORIGIN

```

```

Query Match          8.4%;  Score 36.6;  DB 2;  Length 134580;
Best Local Similarity 53.1%;  Pred. No. 15;
Matches    78;  Conservative    0;  Mismatches    69;  Indels    0;  Gaps    0;
Qy    210 tctttgcgtccgaatctcgagacgagatttttaaggggggggctgtaacacccag 269
| ||||||| | ||| | |||| | ||| ||| | ||| | ||| | ||| | ||| | |
Db 126945 TATTGGCTACAAATGACAGGATTCGTTGTTTATGGCTGAAGAGTATAACACTACAT 126886

Qy    270 gtgttatattctgctcgacaacgagttatggaaattaagcacgttatatcagtgaataaaa 329
| ||| | | ||| | | ||| | ||| | ||| | | | | | | | | | | | | |
Db 126885 ATGTATGAAGTAAGAAAAAGAAAGTGTAGGACTGAGAATGGGCTCTGGATGAAAGGAG 126826

Qy    330 cagataactaaaatttaatcatttcgc 356
| ||| | | ||| | ||| | | |
Db 126825 AAAATAGAAGGCTTCTTCATCTGCC 126799

```

```

RESULT 15
AX083744
LOCUS      AX083744      1141 bp      DNA
DEFINITION Sequence 22 from Patent WO0111061.
ACCESSION  AX083744
VERSION    AX083744.1  GI:13185472
KEYWORDS
SOURCE     synthetic construct.
ORGANISM   synthetic construct
            artificial sequence.
REFERENCE  1 (bases 1 to 1141)
AUTHORS   Kunst,L. and Clemens,S.
TITLE     Regulation of embryonic transcription in plants
JOURNAL   Patent: WO 0111061-A 22 15-FEB-2001;
            UNIVERSITY OF BRITISH COLUMBIA (CA)
FEATURES
source      Location/Qualifiers
            1. .1141
            /organism="synthetic construct"
            /db_xref="taxon:32630"
promoter    1. .1141
            /note="consensus sequence of A.t., L.a., and B.n. FAE1
            promoters"
BASE COUNT   123 a      32 c      42 g      112 t      832 others
ORIGIN

```

Search completed: February 7, 2002, 11:16:33  
Job time: 10519 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 11:01:03 ; Search time 428.31 Seconds  
(without alignments)  
870.716 Million cell updates/sec

Title: US-09-394-745-7826  
Perfect score: 435  
Sequence: 1 aattcacgggcccacgcacg.....cgtccgggctttcctgaat 435  
  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
  
Searched: 930621 seqs 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_1101:  
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:  
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:  
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:  
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:  
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:  
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:  
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:  
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:  
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:  
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:  
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:  
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:  
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:  
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:  
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:  
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:  
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:  
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:  
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:  
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:  
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:  
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query				Description	
		Match	Length	DB	ID		
	1	55.4	12.7	936	22	AAF58252	Oligonucleotide D1
c	2	55.4	12.7	936	22	AAF58252	Oligonucleotide D1
	3	55.4	12.7	936	22	AAF58254	Oligonucleotide D1
c	4	55.4	12.7	936	22	AAF58254	Oligonucleotide D1
	5	55.4	12.7	936	22	AAF58257	Oligonucleotide D1
c	6	55.4	12.7	936	22	AAF58257	Oligonucleotide D1
	7	55.4	12.7	936	22	AAF58259	Oligonucleotide D2
c	8	55.4	12.7	936	22	AAF58259	Oligonucleotide D2
	9	55.4	12.7	936	22	AAF58262	Oligonucleotide D2
c	10	55.4	12.7	936	22	AAF58262	Oligonucleotide D2
	11	55.4	12.7	938	22	AAF58255	Oligonucleotide D1
c	12	55.4	12.7	938	22	AAF58255	Oligonucleotide D1
c	13	35.8	8.2	394	22	AAH55491	Human breast tumou
c	14	35.2	8.1	244	22	AAF58238	Oligonucleotide D1
	15	35.2	8.1	2408	21	AAC76025	Human ORFX ORF1580

16	35.2	8.1	2718	22	AAI60849	Human polynucleoti
17	35.2	8.1	3038	20	AAZ34199	Human PRO541 nucle
18	35.2	8.1	3038	21	AAC78550	Human PRO541 (UNQ3
19	34.6	8.0	244	22	AAF58238	Oligonucleotide D1
20	34.4	7.9	374	21	AAH30824	Human colon cancer
c 21	33.6	7.7	1388	21	AAC47854	Arabidopsis thalia
22	31.8	7.3	2719	20	AAZ33545	Human breast tumou
23	31.8	7.3	3287	20	AAZ41313	Human normal ovari
24	31.8	7.3	6920	22	AAH98318	Chicken EST-derive
25	31.4	7.2	1038	21	AAZ58291	Sorghum resistance
26	31.4	7.2	2271	22	AAI57627	Human colorectal c
27	31.4	7.2	2271	22	AAI57628	Human colorectal c
28	31	7.1	745	21	AAC71947	Single nucleotide
c 29	31	7.1	1902	20	AAX20309	Borrelia burgdorfe
30	30.8	7.1	2892	21	AAZ32971	Human wild-type pr
31	30.8	7.1	2892	21	AAZ38227	Human K309N mutant
c 32	30.8	7.1	68940	20	AAX57351	Human chromosome 6
c 33	30.8	7.1	1082138	21	AAF22305	Arabidopsis thalia
34	30.6	7.0	682	18	AAT84195	DNA encoding a Sta
c 35	30.6	7.0	1356	22	AAS00196	S. aureus DNA enco
36	30.6	7.0	1787	21	AAA38796	K. aestuarii carbo
37	30.6	7.0	2018	21	AAA26865	Essential Staphylo
38	30.6	7.0	2018	22	AAS08016	Staphylococcus aur
39	30.6	7.0	2018	22	AAF91547	Staphylococcus aur
c 40	30.6	7.0	2187	21	AAZ37455	FemR315 coding seq
41	30.6	7.0	4291	20	AAV69278	Human T85 cDNA. H
c 42	30.6	7.0	8155	18	AAV74374	Staphylococcus aur
43	30.6	7.0	12283	12	AAQ10353	Nucleotide sequenc
44	30.2	6.9	2849	20	AAX20293	Borrelia burgdorfe
c 45	30	6.9	3247	21	AAA46567	DNA encoding the p

#### ALIGNMENTS

RESULT 1  
 AAF58252  
 ID AAF58252 standard; DNA; 936 BP.  
 XX  
 AC AAF58252;  
 XX  
 DT 24-APR-2001 (first entry)  
 XX  
 DE Oligonucleotide D1835.  
 XX  
 KW Electron-transfer group; ETM; mismatch; genotyping;  
 KW gene expression; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200107665-A2.  
 XX  
 PD 01-FEB-2001.  
 XX  
 PF 26-JUL-2000; 2000WO-US20476.  
 XX  
 PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.  
XX  
PA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX  
PI Umek RM;  
XX  
DR WPI; 2001-159728/16.  
XX  
PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface -  
XX  
PS Example 6; Page 127; 159pp; English.  
XX  
CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX  
SO Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

RESULT 2  
AAF58252/c  
ID AAF58252 standard; DNA; 936 BP.  
XX  
AC AAF58252;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Oligonucleotide D1835.  
XX  
KW Electron-transfer group; ETM; mismatch; genotyping;  
KW gene expression; ss.  
XX  
OS Synthetic.  
XX  
PN WO200107665-A2.  
XX  
PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000WO-US20476.  
XX  
PR 26-JUL-1999; 99US-0145695.  
PR 17-MAR-2000; 2000US-0190259.  
XX  
PA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX  
PI Umek RM;  
XX  
DR WPI; 2001-159728/16.  
XX  
PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface -  
XX  
PS Example 6; Page 127; 159pp; English.  
XX  
CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX  
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

RESULT 3  
AAF58254  
ID AAF58254 standard; DNA; 936 BP.  
XX  
AC AAF58254;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Oligonucleotide D1875.  
XX  
KW Electron-transfer group; ETM; mismatch; genotyping;  
KW gene expression; ss.  
XX  
OS Synthetic.  
XX  
PN WO200107665-A2.  
XX  
PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000WO-US20476.  
XX  
PR 26-JUL-1999; 99US-0145695.  
PR 17-MAR-2000; 2000US-0190259.  
XX  
PA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX  
PI Umek RM;  
XX  
DR WPI; 2001-159728/16.  
XX  
PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface -  
XX  
PS Example 6; Page 127; 159pp; English.

XX  
CC The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.  
XX  
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 12.7%; Score 55.4; DB 22; Length 936;  
Best Local Similarity 1.3%; Pred. No. 7.4e-08;  
Matches 5; Conservative 231; Mismatches 147; Indels 0; Gaps 0;  
  
Qy 33 ttatgtgttcttctggcagacatcgccctattggggacatctctaaataggcttaagg 92  
Db 367 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 426  
  
Qy 93 cgatacatgttatgtccactagagaaacaacatcctgagacactcacctttatggaaa 152  
Db 427 ww 486  
  
Qy 153 tgtctcgcgattatcgctgatgtggacatgttacatgctctactctaaaagtct 212  
Db 487 wwwwwwcww 546  
  
Qy 213 tttgctccgaatctcgagacgagatttttaaggggggggctgtaacaccccaggtg 272  
Db 547 ww 606  
  
Qy 273 ttatattctgctcgacaacgagtatggaattaaggcacgttatcatgtaatgaaacag 332  
Db 607 ww 666  
  
Qy 333 atactaaaatttaatcatttcgctatcgcgattttatcatgtatctgtccatctg 392  
Db 667 wwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 726  
  
Qy 393 gtgagtgtgacatcattttatt 415  
Db 727 wwwwwwwwwwwwwwwwwwwwww 749

RESULT 4  
AAF58254/c  
ID AAF58254 standard; DNA; 936 BP.  
XX  
AC AAF58254;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Oligonucleotide D1875.  
XX  
KW Electron-transfer group; ETM; mismatch; genotyping;  
KW gene expression; ss.  
XX

OS Synthetic.  
XX  
PN WO200107665-A2.  
XX  
PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000WO-US20476.  
XX  
PR 26-JUL-1999; 99US-0145695.  
PR 17-MAR-2000; 2000US-0190259.  
XX  
PA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX  
PI Umek RM;  
XX  
DR WPI; 2001-159728/16.  
XX  
PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface -  
XX  
PS Example 6; Page 127; 159pp; English.  
XX  
CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX  
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

RESULT 5  
AAF58257  
ID AAF58257 standard; DNA; 936 BP.  
XX  
AC AAF58257;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Oligonucleotide D1954.  
XX  
KW Electron-transfer group; ETM; mismatch; genotyping;  
KW gene expression; ss.  
XX  
OS Synthetic.  
XX  
PN WO200107665-A2.  
XX  
PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000WO-US20476.  
XX  
PR 26-JUL-1999; 99US-0145695.  
PR 17-MAR-2000; 2000US-0190259.  
XX  
PA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX  
PI Umek RM;  
XX  
DR WPI; 2001-159728/16.  
XX  
PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface -  
XX  
PS Example 6; Page 127; 159pp; English.  
XX  
CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX  
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 12.7%; Score 55.4; DB 22; Length 936;  
Best Local Similarity 1.3%; Pred. No. 7.4e-08;

RESULT 6  
AAF58257/c  
ID AAF58257 standard; DNA; 936 BP.  
XX  
AC AAF58257;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Oligonucleotide D1954.  
XX  
KW Electron-transfer group; ETM; mismatch; genotyping;  
KW gene expression; ss.  
XX  
OS Synthetic.  
XX  
PN WO200107665-A2.  
XX  
PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000WO-US20476.  
XX  
PR 26-JUL-1999; 99US-0145695.  
PR 17-MAR-2000; 2000US-0190259.  
XX  
PA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX

PI Umek RM;  
XX  
DR WPI; 2001-159728/16.  
XX  
PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface -  
XX  
PS Example 6; Page 127; 159pp; English.  
XX  
CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX  
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 12.7%; Score 55.4; DB 22; Length 936;  
Best Local Similarity 1.3%; Pred. No. 7.4e-08;  
Matches 5; Conservative 231; Mismatches 147; Indels 0; Gaps 0;

RESULT 7  
AAF58259  
ID AAF58259 standard; DNA; 936 BP.  
XX

AC AAF58259;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Oligonucleotide D2004.  
XX  
KW Electron-transfer group; ETM; mismatch; genotyping;  
KW gene expression; ss.  
XX  
OS Synthetic.  
XX  
PN WO200107665-A2.  
XX  
PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000WO-US20476.  
XX  
PR 26-JUL-1999; 99US-0145695.  
PR 17-MAR-2000; 2000US-0190259.  
XX  
PA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX  
PI Umek RM;  
XX  
DR WPI; 2001-159728/16.  
XX  
PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface -  
XX  
PS Example 6; Page 128; 159pp; English.  
XX  
CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX  
SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 12.7%; Score 55.4; DB 22; Length 936;  
Best Local Similarity 1.3%; Pred. No. 7.4e-08;  
Matches 5; Conservative 231; Mismatches 147; Indels 0; Gaps 0;

Qy 93 cgatacatgttatgtccactagagaacaacatcctgagacactcaccttatttgaaa 152  
      ::: :: :::: : : :: : :::: :: : :: : : : : : : : : ::::::: :::

Qy 153 tgtctcgcgattatcgctgatgtggacatgtgttacatgcttctactcttaaaagtct 212

Qy 213 tttgctccaatctcgagacgagattatTTAAGGGGGAGGGCTGTAACACCCAGGTG 272  
      ::: :    ::: :    : :    : ::::::::::::    :    : |::: :    : :  
Db 547 WWW 606

Qy 273 ttatattctgctcacaacgagtatggattaAGCACGTTATATCAGTGAATGAAACAG 332  
      :::::::::: :    : :    : :::    : :::::::    :    : ::::: :  
Db 607 WWW 666

Qy 333 atactaaaatttaatcatTTCGCTATCGCGATTATATCGTATCTGTTCCATCTGTC 392  
      ::: ::::::::::: :::::    |:::    ::::::::::: ::: : ::: :: :  
Db 667 WWW 726

Qy 393 gtgagtgtgacatcatTTTATT 415  
      : : : :    : :: :::::::  
Db 727 WWWWWWWWWWWWWWWWWWWWWWWWWWWWW 749

RESULT 8  
AAF58259/c  
ID AAF58259 standard; DNA; 936 BP.  
XX  
AC AAF58259;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Oligonucleotide D2004.  
XX  
KW Electron-transfer group; ETM; mismatch; genotyping;  
KW gene expression; ss.  
XX  
OS Synthetic.  
XX  
PN WO200107665-A2.  
XX  
PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000WO-US20476.  
XX  
PR 26-JUL-1999; 99US-0145695.  
PR 17-MAR-2000; 2000US-0190259.  
XX  
PA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX  
PI Umek RM;  
XX  
DR WPI; 2001-159728/16.  
XX  
PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface -  
XX  
PS Example 6; Page 128; 159pp; English.  
XX  
CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic

RESULT 9  
AAF58262  
ID AAF58262 standard; DNA; 936 BP.  
XX  
AC AAF58262;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Oligonucleotide D2007.  
XX  
KW Electron-transfer group; ETM; mismatch; genotyping;  
KW gene expression; ss.  
XX  
OS Synthetic.  
XX  
PN WO200107665-A2.  
XX

PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000WO-US20476.  
XX  
PR 26-JUL-1999; 99US-0145695.  
PR 17-MAR-2000; 2000US-0190259.  
XX  
PA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX  
PI Umek RM;  
XX  
DR WPI; 2001-159728/16.  
XX  
PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface -  
XX  
PS Example 6; Page 128; 159pp; English.  
XX  
CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX  
SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

Query Match 12.7%; Score 55.4; DB 22; Length 936;  
Best Local Similarity 1.3%; Pred. No. 7.4e-08;  
Matches 5; Conservative 231; Mismatches 147; Indels 0; Gaps 0;

Qy 213 tttgctccgaatctcgagacgagattatTTAAGGGGGGAGGGCTGTaacacCCCAggTG 272  
      ::: :   ::: :   : :   : :::::::::::::   :    ::: :   :    : :  
Ei 517

Qy 333 atactaaaatthaatcatttcgcgtatcgcatatcgatctgttccatctgtc 392

Qy 393 gtgagtgtgacatcattttatt 415  
: : : : : : : : : : : :  
Db 727 wwwwwwwwwwwwwwwwwwwwwwww 749

RESULT 10  
AAF58262/c  
ID AAF58262 standard; DNA; 936 BP.  
XX  
AC AAF58262;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Oligonucleotide D2007.  
XX  
KW Electron-transfer group; ETM; mismatch; genotyping;  
KW gene expression; ss.  
XX  
OS Synthetic.  
XX  
PN WO200107665-A2.  
XX  
PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000WO-US20476.  
XX  
PR 26-JUL-1999; 99US-0145695.  
PR 17-MAR-2000; 2000US-0190259.  
XX  
PA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX  
PI Umek RM;  
XX  
DR WPI; 2001-159728/16.  
XX  
PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface -  
XX  
PS Example 6; Page 128; 159pp; English.  
XX  
CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX  
SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

Query Match 12.7%; Score 55.4; DB 22; Length 936;  
Best Local Similarity 1.3%; Pred. No. 7.4e-08;  
Matches 5; Conservative 231; Mismatches 147; Indels 0; Gaps 0;

Qy 33 ttatgtgttcttctggcagacatgcgcctctattggtgacatctctaaatttagcttaagg 92  
:::: : :: : : : : : : : : : : : : : : : : : :

RESULT 11  
AAF58255  
ID AAF58255 standard; DNA; 938 BP.  
XX  
AC AAF58255;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Oligonucleotide D1876.  
XX  
KW Electron-transfer group; ETM; mismatch; genotyping;  
KW gene expression; ss.  
XX  
OS Synthetic.  
XX  
PN WO200107665-A2.  
XX  
PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000WO-US20476.  
XX  
PR 26-JUL-1999; 99US-0145695.  
PR 17-MAR-2000; 2000US-0190259.  
XX  
PA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX  
PI Umek RM;  
XX  
DR WPI; 2001-159728/16.  
XX

PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface -

PS Example 6: Page 127: 159pp: English.

13  
xx

CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.

xx

SO Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match 12.7%; Score 55.4; DB 22; Length 938;  
Best Local Similarity 1.3%; Pred. No. 7.4e-08;  
Matches 5; Conservative 231; Mismatches 147; Indels 0; Gaps 0;

Qy 33 ttatgtgttcttctggcagacatgcctctattggatggacatctctaaatttagcttaagg 92  
::::: :::: : :: : || :: : :: : :: :: :: :: :: :: :: :: :: :: :: :: :: ::

Qy 93 cgatacatgttatgtccactagagaacaacatcctgagacactcaccttatttgaaa 152  
      ::: :: :::: :   : :: : :::: :: : ::   : : : :: : : :   : ::::::: :::

RESULT 12  
AAF58255/c

ID AAF58255  
MM

xx

AC AAF 58233,  
XX  
DT 24-APR-2001 (first entry)

DE Oligonucleotide D1876.  
XX  
KW Electron-transfer group; ETM; mismatch; genotyping;  
KW gene expression; ss.  
XX  
OS Synthetic.  
XX  
PN WO200107665-A2.  
XX  
PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000WO-US20476.  
XX  
PR 26-JUL-1999; 99US-0145695.  
PR 17-MAR-2000; 2000US-0190259.  
XX  
PA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX  
PI Umek RM;  
XX  
DR WPI; 2001-159728/16.  
XX  
PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface -  
XX  
PS Example 6; Page 127; 159pp; English.  
XX  
CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX  
SO Sequence 938 BP: 4 A: 144 C: 9 G: 5 T: 776 other:

Query Match 12.7%; Score 55.4; DB 22; Length 938;  
 Best Local Similarity 1.3%; Pred. No. 7.4e-08;  
 Matches 5; Conservative 231; Mismatches 147; Indels 0; Gaps 0;

RESULT 13  
AAH55491/c  
ID AAH55491 standard; DNA; 394 BP.  
XX  
AC AAH55491;  
XX  
DT 04-SEP-2001 (first entry)  
XX  
DE Human breast tumour protein clone 26664 DNA sequence.  
XX  
KW Cytostatic; vaccine; human; breast tumour protein; breast cancer;  
KW gene therapy; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200140269-A2.  
XX  
PD 07-JUN-2001.  
XX  
PF 29-NOV-2000; 2000WO-US32520.  
XX  
PR 30-NOV-1999; 99US-0451651.  
PR 22-FEB-2000; 2000US-0510662.  
PR 10-MAR-2000; 2000US-0523586.  
PR 07-APR-2000; 2000US-0545068.  
PR 15-MAY-2000; 2000US-0571025.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Dillon DC, Day CH, Jiang Y, Houghton RL, Mitcham JL, Wang A;  
XX  
DR WPI; 2001-356154/37.  
XX  
PT Breast tumor polypeptides and the nucleic acids that encode them,  
PT useful for the prevention, diagnosis and treatment of breast cancer -  
XX  
PS Claim 5; Page 125; 221pp; English.  
XX  
CC The present sequence is a human breast tumour protein coding sequence.  
CC This sequence may be used in the prevention, diagnosis and treatment of  
CC diseases associated with inappropriate expression of the breast tumour  
CC protein e.g. breast cancer. For example, this sequence may be used to  
CC treat disorders associated with decreased expression by rectifying

CC mutations or deletions in a patient's genome that affect the activity of  
CC breast tumour protein by expressing inactive proteins or to supplement  
CC the patients own production of the breast tumour protein. Additionally,  
CC the present sequence may be used to produce the breast tumour protein, by  
CC inserting the nucleic acids into a host cell and culturing the cell to  
CC express the protein. The present sequence and its complementary sequences  
CC may also be used as DNA probes in diagnostic assays to detect and  
CC quantitate the presence of similar nucleic acids in samples, and  
CC therefore which patients may be in need of restorative therapy.  
XX  
SQ Sequence 394 BP; 133 A; 69 C; 47 G; 140 T; 5 other;

Query Match 8.2%; Score 35.8; DB 22; Length 394;  
Best Local Similarity 57.1%; Pred. No. 0.098;  
Matches 64; Conservative 0; Mismatches 48; Indels 0; Gaps 0;  
Qy 264 ccccaggtttatattctgctcgacaacgagtatgaaattaagcacgttatcatcagtga 323  
|| || | | || || || | || || | || || || || | | || || || || || || ||  
Db 226 CCAAATGCATATAAATCTTGATAAACAAAGTNTATAAAATAAAACATGGGACATTAGCTT 167  
Qy 324 atgaaaacagataactaaaatttaatcatttcgctatcgcgatttttatatcg 375  
| | | | | | || || | || || | | || | | || | | || | | || |  
Db 166 TGGGAAAAGTAATGAAAATATAATGGTTTAGAAATCCTGTAAATATTG 115

RESULT 14  
AAF58238/c  
ID AAF58238 standard; DNA; 244 BP.  
XX  
AC AAF58238;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Oligonucleotide D1250:D1102.  
XX  
KW Electron-transfer group; ETM; mismatch; genotyping;  
KW gene expression; ss.  
XX  
OS Synthetic.  
XX  
PN WO200107665-A2.  
XX  
PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000WO-US20476.  
XX  
PR 26-JUL-1999; 99US-0145695.  
PR 17-MAR-2000; 2000US-0190259.  
XX  
PA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX  
PI Umek RM;  
XX  
DR WPI; 2001-159728/16.  
XX  
PT Nucleic acids containing electron-transfer group, useful as labels in

PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface -

xx

PS Example 4; Page 120; 159pp; English.

xx

CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.

xx

SQ Sequence 244 BP; 19 A; 9 C; 12 G; 10 T; 194 other;

Query Match 8.1%; Score 35.2; DB 22; Length 244;  
Best Local Similarity 6.0%; Pred. No. 0.13;  
Matches 10; Conservative 99; Mismatches 57; Indels 0; Gaps 0;

Qy 250 ggagggtgtaacaccccccaggtgttatattctgctcgacaacgagtgatggaaattaagca 309

Qy 310 cgttatatcagtgaatgaaacagatactaaaatttaatcatttcgctatcgcgattttt 369

156 УЧЕБНИК ПО АВТОМАТИЗАЦИИ И СИСТЕМЫ УПРАВЛЕНИЯ 87

Qy 370 atatcgatatctgttccatctgtcgtagtgtgacatcattttatt 415  
..... .

RESULT 15

AAC76025

ID: AAC76025 standard; cDNA; 2408 BP.

xx

AC AAC76025;

xx

DT 08-FEB-2001 (first entry)

xx

DE Human ORFX ORF1580 polynucleotide sequence SEQ ID NO: 3159.

xx

KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive; ss.

xx

OS Homo sapiens.  
XX  
PN WO200058473-A2..  
XX  
PD 05-OCT-2000.  
XX  
PF 31-MAR-2000; 2000WO-US08621.  
XX  
PR 31-MAR-1999; 99US-0127607.  
PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX  
PA (CURA<sup>®</sup>) CURAGEN CORP.  
XX  
PI Shimkets RA, Leach M;  
XX  
DR WPI; 2000-602362/57.  
DR P-PSDB; AAB41816.  
XX  
PT Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
PS Claim 5; Page 2375-2377; 5507pp; English.  
XX  
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;  
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antithyroid; and antianaemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
SQ Sequence 2408 BP; 698 A; 516 C; 567 G; 625 T; 2 other;

Query Match 8.1%; Score 35.2; DB 21; Length 2408;  
 Best Local Similarity 57.1%; Pred. No. 0.32;  
 Matches 64; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

. Qy 264 cccagggtttatattctgctcgacaacgagtatggattaagcacgttatatcagtga 323

Db 2078 ccaaatgcataaaatcttataaacaaggctataaaaataacatgggacattagtt 2137  
Qy 324 atgaaacagataactaaaatttaatcatttcgctatcgcatatcg 375  
| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 2138 tggaaaagtaatgaaaatataatggtttagaaatccgtgttaaatattg 2189

Search completed: February 7, 2002, 11:01:06  
Job time: 5052 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 11:42:57 ; Search time 172.96 Seconds  
(without alignments)  
569.599 Million cell updates/sec

Title: US-09-394-745-7826  
Perfect score: 435  
Sequence: 1 aattcacgggccgacgcacg.....cgtccggctttcctgaat 435

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query				Description
		Match	Length	DB	ID	
1	30.8	7.1	2892	2	US-08-874-186-44	Sequence 44, Appl

2	30.6	7.0	2018	3	US-08-714-918-16	Sequence 16, Appl
3	30.6	7.0	2018	4	US-09-265-315-16	Sequence 16, Appl
4	30.6	7.0	2018	4	US-09-265-315-16	Sequence 16, Appl
5	30.6	7.0	2018	4	US-09-266-417-16	Sequence 16, Appl
c 6	30.6	7.0	2187	2	US-08-679-635A-1	Sequence 1, Appl
c 7	29.6	6.8	5630	2	US-08-937-931-1	Sequence 1, Appl
c 8	29.6	6.8	5630	4	US-09-285-502-1	Sequence 1, Appl
9	29.4	6.8	2430	4	US-08-845-258-3	Sequence 3, Appl
c 10	29.4	6.8	2430	4	US-08-845-258-40	Sequence 40, Appl
11	29.4	6.8	2430	4	US-08-990-571-3	Sequence 3, Appl
c 12	29.4	6.8	2430	4	US-08-990-571-40	Sequence 40, Appl
13	29.4	6.8	2430	4	US-08-723-142A-3	Sequence 3, Appl
c 14	29.4	6.8	2430	4	US-08-723-142A-40	Sequence 40, Appl
c 15	29.4	6.8	2790	3	US-08-895-601-2	Sequence 2, Appl
16	29.4	6.8	3207	1	US-08-162-081B-35	Sequence 35, Appl
17	29.4	6.8	3207	2	US-08-780-872-35	Sequence 35, Appl
18	29.4	6.8	3207	4	US-09-085-957-35	Sequence 35, Appl
19	29.4	6.8	3240	1	US-08-162-081B-34	Sequence 34, Appl
20	29.4	6.8	3240	2	US-08-780-872-34	Sequence 34, Appl
21	29.4	6.8	3240	4	US-09-085-957-34	Sequence 34, Appl
22	29.4	6.8	3412	1	US-08-162-081B-32	Sequence 32, Appl
23	29.4	6.8	3412	2	US-08-780-872-32	Sequence 32, Appl
24	29.4	6.8	3412	4	US-09-085-957-32	Sequence 32, Appl
c 25	29	6.7	3726	1	US-08-173-497-1	Sequence 1, Appl
c 26	29	6.7	3726	1	US-08-286-889-1	Sequence 1, Appl
c 27	29	6.7	3726	1	US-08-485-618-1	Sequence 1, Appl
c 28	29	6.7	3726	1	US-08-362-652-1	Sequence 1, Appl
c 29	29	6.7	3726	1	US-08-605-672-1	Sequence 1, Appl
c 30	29	6.7	3726	2	US-08-482-293A-1	Sequence 1, Appl
c 31	29	6.7	3726	2	US-08-943-363-1	Sequence 1, Appl
c 32	29	6.7	3726	4	US-09-193-043-1	Sequence 1, Appl
c 33	29	6.7	3785	1	US-08-485-618-98	Sequence 98, Appl
c 34	29	6.7	3785	1	US-08-605-672-98	Sequence 98, Appl
c 35	29	6.7	3785	2	US-08-482-293A-98	Sequence 98, Appl
c 36	29	6.7	3785	2	US-08-943-363-98	Sequence 98, Appl
c 37	29	6.7	3785	4	US-09-193-043-98	Sequence 98, Appl
c 38	29	6.7	3956	1	US-08-485-618-97	Sequence 97, Appl
c 39	29	6.7	3956	1	US-08-605-672-97	Sequence 97, Appl
c 40	29	6.7	3956	2	US-08-482-293A-97	Sequence 97, Appl
c 41	29	6.7	3956	2	US-08-943-363-97	Sequence 97, Appl
c 42	29	6.7	3956	4	US-09-193-043-97	Sequence 97, Appl
43	28.8	6.6	765	2	US-07-690-192-3	Sequence 3, Appl
c 44	28.4	6.5	1379	1	US-08-500-611-1	Sequence 1, Appl
c 45	28.4	6.5	1379	2	US-08-500-694-1	Sequence 1, Appl

#### ALIGNMENTS

RESULT 1  
 US-08-874-186-44  
 ; Sequence 44, Application US/08874186  
 ; Patent No. 5989885  
 ; GENERAL INFORMATION:  
 ;     APPLICANT: Teng, David H-F.  
 ;     APPLICANT: Tavtigian, Sean V.  
 ;     APPLICANT: Perry III, William L.

; APPLICANT: Skolnick, Mark H.  
; TITLE OF INVENTION: SPECIFIC MUTATIONS OF MAP KINASE KINASE  
; TITLE OF INVENTION: 4 (MKK4) IN HUMAN TUMOR CELL LINES IDENTIFY IT AS A  
TUMOR  
; TITLE OF INVENTION: SUPPRESSOR IN VARIOUS TYPES OF CANCER  
; NUMBER OF SEQUENCES: 96  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
; STREET: 1201 New York Avenue, N.W., Suite 1000  
; CITY: Washington  
; STATE: DC  
; COUNTRY: U.S.A.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/874,186  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/782,482  
; FILING DATE: 10-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saxe, Stephen A.  
; REGISTRATION NUMBER: 38,609  
; REFERENCE/DOCKET NUMBER: 24884-121392-01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-962-4848  
; TELEFAX: 202-962-8300  
; INFORMATION FOR SEQ ID NO: 44:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2892 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: 1..1030  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 1031..1179  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: 1180..2892  
US-08-874-186-44

Query Match 7.1%; Score 30.8; DB 2; Length 2892;  
Best Local Similarity 54.4%; Pred. No. 2.5;  
Matches 62; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 302 attaagcacgttatatcagtgaatgaaacagatactaaaattaatcatttcgctatcg 361  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 2272 AATTCACACAGTATGGATAGTTATATAATTGCATAATGTGATCATTATGTATT 2331  
  
Qy 362 cgattttatatcgatctgtccatctgcgtgagggtgacatcattttatt 415  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 2332 CATTTTATGACATATTGCTAAAATGATCTGTGTAAGTCATAGGGTATAAT 2385

RESULT 2

US-08-714-918-16

; Sequence 16, Application US/08714918

; Patent No. 6037123

; GENERAL INFORMATION:

; APPLICANT: Benton, Bret

; APPLICANT: Lee, Ving

; APPLICANT: Malouin, Francois

; APPLICANT: Martin, Patrick K.

; APPLICANT: Schmid, Molly B.

; APPLICANT: Sun, Dongxu

; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS ANTIBACTERIAL

; TITLE OF INVENTION: TARGET GENES

; NUMBER OF SEQUENCES: 111

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lyon & Lyon

; STREET: 633 West Fifth Street

; STREET: Suite 4700

; CITY: Los Angeles

; STATE: California

; COUNTRY: U.S.A.

; ZIP: 90071-2066

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; MEDIUM TYPE: storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: IBM P.C. DOS 5.0

; SOFTWARE: Word Perfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/714,918

; FILING DATE: September 13, 1996

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/009,102

; FILING DATE: December 22, 1995

; APPLICATION NUMBER: 60/003,798

; FILING DATE: September 15, 1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Warburg, Richard J.

; REGISTRATION NUMBER: 32,327

; REFERENCE/DOCKET NUMBER: 222/005

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (213) 489-1600

; TELEFAX: (213) 955-0440

; TELEX: 67-3510

; INFORMATION FOR SEQ ID NO: 16:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2018 base pairs

; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-714-918-16

Query Match 7.0%; Score 30.6; DB 3; Length 2018;  
Best Local Similarity 56.4%; Pred. No. 2.5;  
Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;  
  
Qy 329 acagataactaaaatttaatcatttcgctatcgcgattttatatcgatctgttccatc 388  
| | | | | || | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1444 ATAACTTCTTAACGTCAACATTTCTTCAACACGATATTATCTGTTACCGGTACGTTA 1503  
  
Qy 389 tgtcgtgagtgtgacatcattttattcgccggcttc 429  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1504 ATTAATGATTGTGGATATTTTCATTTGCCAGCTAAC 1544

RESULT 3  
US-09-265-315-16  
; Sequence 16, Application US/09265315  
; Patent No. 6187541  
; GENERAL INFORMATION:  
; APPLICANT: Benton, Bret  
; APPLICANT: Lee, Ving J.  
; APPLICANT: Malouin, Francois  
; APPLICANT: Martin, Patrick K.  
; APPLICANT: Schmid, Molly B.  
; APPLICANT: Sun, Dongxu  
; TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS  
; TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS  
; TITLE OF INVENTION: TARGET GENES  
; NUMBER OF SEQUENCES: 111  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/265,315  
; FILING DATE: March 9, 1999  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/714,918  
; FILING DATE: September 13, 1996  
; APPLICATION NUMBER: 60/009,102  
; FILING DATE: December 22, 1995

; APPLICATION NUMBER: 60/003,798  
; FILING DATE: September 15, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 240/247  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2018 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-265-315-16

Query Match 7.0%; Score 30.6; DB 4; Length 2018;  
Best Local Similarity 56.4%; Pred. No. 2.5;  
Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;  
  
Qy 329 acagataactaaaatattaatcatttcgctatcgcgattttatatcgatctgttccatc 388  
| | | || | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1444 ATAACTTCTTAACGTCAACATTTCTTCAACACGATATTATCTGTTACCGTACGTTA 1503  
  
Qy 389 tgcgtgagtgtgacatcatttattcgccggcttc 429  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1504 ATTAATGATTGTGGATATTTTCATTGTCCAGCTAAC 1544

RESULT 4  
US-09-265-315-16  
; Sequence 16, Application US/09265315  
; Patent No. 6187541  
; GENERAL INFORMATION:  
; APPLICANT: Benton, Bret  
; APPLICANT: Lee, Ving J.  
; APPLICANT: Malouin, Francois  
; APPLICANT: Martin, Patrick K.  
; APPLICANT: Schmid, Molly B.  
; APPLICANT: Sun, Dongxu  
; TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS  
; TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS  
; TITLE OF INVENTION: TARGET GENES  
; NUMBER OF SEQUENCES: 111  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/265,315  
; FILING DATE: March 9, 1999  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/714,918  
; FILING DATE: September 13, 1996  
; APPLICATION NUMBER: 60/009,102  
; FILING DATE: December 22, 1995  
; APPLICATION NUMBER: 60/003,798  
; FILING DATE: September 15, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 240/247  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2018 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-09-265-315-16

Query Match 7.0%; Score 30.6; DB 4; Length 2018;  
Best Local Similarity 56.4%; Pred. No. 2.5;  
Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 329 acagataactaaaatttaatcatttcgctatcgcgattttatatcgatctgtccatc 388  
| | | | || || || || || || || || || || || || || || || || || || || |  
Db 1444 ATAACTTCTTAACGTCAACATTTCTTCAACACGATATTATCTGTTACCGTACGTTA 1503  
  
Qy 389 tgtcgtgagtgtgacatcattttattcgtccgggccttc 429  
| ||| |||| | ||| ||| ||| ||| | |||  
Db 1504 ATTAATGATTGTGGATATTTTCATTTGCCAGCTAAC 1544

RESULT 5

US-09-266-417-16

; Sequence 16, Application US/09266417  
; Patent No. 6228588  
; GENERAL INFORMATION:  
; APPLICANT: Benton, Bret  
; APPLICANT: Lee, Ving J.  
; APPLICANT: Malouin, Francois  
; APPLICANT: Martin, Patrick K.  
; APPLICANT: Schmid, Molly B.  
; APPLICANT: Sun, Dongxu  
; TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS

; TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS  
; TITLE OF INVENTION: TARGET GENES  
; NUMBER OF SEQUENCES: 111  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/266,417  
; FILING DATE: March 9, 1999  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/714,918  
; FILING DATE: September 13, 1996  
; APPLICATION NUMBER: 60/009,102  
; FILING DATE: December 22, 1995  
; APPLICATION NUMBER: 60/003,798  
; FILING DATE: September 15, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 240/248  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2018 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-09-266-417-16

Query Match 7.0%; Score 30.6; DB 4; Length 2018;  
Best Local Similarity 56.4%; Pred. No. 2.5;  
Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 329 acagataactaaaatttaatcatttcgctatcgcgattttatatcgatctgttccatc 388  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 1444 ATAACTTCTTAACGTCAACATTTCTTCAACACGATATTATCTGTTACCGTACGTTA 1503

Qy 389 tgcgtgagtgtgacatcatttattcgccggcttc 429  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 1504 ATTAATGATTGTGGATATTTTCATTTGTCCAGCTAAC 1544

RESULT 6  
US-08-679-635A-1/c  
; Sequence 1, Application US/08679635A  
; Patent No. 5985643  
; GENERAL INFORMATION:  
; APPLICANT: Tomasz, Alexander  
; APPLICANT: Delencastre, Herminia  
; TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF  
; TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David A. Jackson, Esq.  
; STREET: 411 Hackensack Ave, Continental Plaza, 4th  
; STREET: Floor  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/679,635A  
; FILING DATE: 10-JUL-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-141  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2187 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Staphylococcus aureus  
; STRAIN: RUSA 315  
US-08-679-635A-1

Query Match 7.0%; Score 30.6; DB 2; Length 2187;  
Best Local Similarity 56.4%; Pred. No. 2.6;  
Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 329 acagataactaaaatttaatcatttcgctatcgcgattttatatcgatctgttccatc 388  
| | | | || | | | | || | | | | | | | | | | | | | | | | | | | | | |  
Db 1627 ATAACCTCTTAAACGTCAACATTTCTTCAACACGATATTATCTGTTACCGTACGTTA 1568

Qy 389 tgtcgtgagtgtgacatcattttattcgccggcttc 429  
| ||| |||| | |||| ||| ||| | |||  
Db 1567 ATTAATGATTGTGGATATTTTCATTTGCCAGCTAATTC 1527

RESULT 7  
US-08-937-931-1/c  
; Sequence 1, Application US/08937931  
; Patent No. 5935792  
; GENERAL INFORMATION:  
; APPLICANT: Rubin, Gerald M.  
; APPLICANT: Pan, Duoja  
; APPLICANT: Rooke, Jenny  
; APPLICANT: Yavari, Reza  
; APPLICANT: Xu, Tian  
; TITLE OF INVENTION: KUZ: A No. 5935792el Family of Metalloproteases  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 BUSH STREET, SUITE 3200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/937,931  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, RICHARD A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: B97-081  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 343-4341  
; TELEFAX: (415) 343-4342  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5630 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-937-931-1

Query Match 6.8%; Score 29.6; DB 2; Length 5630;  
Best Local Similarity 54.6%; Pred. No. 7.4;  
Matches 59; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 273 tttatattctgctcgacaacgagtatggaattaaggcacgttatcatgtaatgaaacag 332  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 5209 TTACTTTATTATTCTATAATTGCATTCTGATTTCAATTAAATTAGTTAATCAAAAAT 5150

Qy 333 atactaaaatthaatcatttcgctatcgcgattttatatcgatct 380  
|| | | | ||| | | |||| | ||| | | | | | | | | | | | |  
Db 5149 ATCATTCACTTCAGCTTTCTGTATTACGAAATTGTCTCCTTTT 5102

RESULT 8  
US-09-285-502-1/c  
; Sequence 1, Application US/09285502  
; Patent No. 6190876  
; GENERAL INFORMATION:  
; APPLICANT: Rubin, Gerald M.  
; APPLICANT: Pan, Duoja  
; APPLICANT: Rooke, Jenny  
; APPLICANT: Yavari, Reza  
; APPLICANT: Xu, Tian  
; TITLE OF INVENTION: KUZ: A No. 6190876el Family of Metalloproteases  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 BUSH STREET, SUITE 3200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/285,502  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/937,931  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, RICHARD A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: B97-081  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 343-4341  
; TELEFAX: (415) 343-4342  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5630 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-09-285-502-1

Query Match 6.8%; Score 29.6; DB 4; Length 5630;  
Best Local Similarity 54.6%; Pred. No. 7.4;  
Matches 59; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 273 tttatattctgctgacaacgagtatggattaaggcacgttatcatcgaaatgaaacag 332  
|| | || | | | || | || | || | || | || | || | || |  
Db 5209 TTACTTTATTATTCTATAATTGCATTCGTATTCATTTAATTAGTTAATCAAAAAT 5150

Qy 333 atactaaaatttaatcatttcgcatacgatttatcatcgatct 380  
|| | | | || | | || || | || | || | || | | | |  
Db 5149 ATCATTCACTTCAGTTCTGTATTACGAAATTGTCTCCTTTT 5102

RESULT 9  
US-08-845-258-3  
; Sequence 3, Application US/08845258  
; Patent No. 6183976  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Sleath, Paul R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/845,258  
; FILING DATE: 24-APR-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.426C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2430 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-845-258-3

Query Match 6.8%; Score 29.4; DB 4; Length 2430;  
Best Local Similarity 51.1%; Pred. No. 6.3;  
Matches 69; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 274 ttatattctgctcgacaacgagtatggattaaggcacgttatcagtgaatgaaacaga 333  
| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1385 TAATAAAATTAGTATACAATGATTATATTACAGATGACTATTGATTATTGTATCAATTAAA 1444

Qy 334 tactaaaatttaatcatttcgctatcgcgattttatatcgtatctgttcatctgtcg 393  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1445 TATTGATTATTAATGATATCATATATGTATATGTTAATGATTGATTGTTACGTTGTG 1504

Qy 394 tgagtgtgacatcat 408  
| | | | | | | |  
Db 1505 AATATGTTATATAAT 1519

RESULT 10  
US-08-845-258-40/c  
; Sequence 40, Application US/08845258  
; Patent No. 6183976  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Sleath, Paul R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/845,258  
; FILING DATE: 24-APR-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.426C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2430 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-08-845-258-40

Query Match                    6.8%; Score 29.4; DB 4; Length 2430;  
 Best Local Similarity    51.1%; Pred. No. 6.3;  
 Matches    69; Conservative    0; Mismatches    66; Indels    0; Gaps    0;  
  
 Qy    274 ttatattctgctcgacaacgagtatggaattaaggcacgttatcagtgaatgaaacaga 333  
       | | || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 Db    1046 TAATAAATTAGTATACAATGATTATATTACAGATGACTATTGATTATTGTATCAATTAAA 987  
  
 Qy    334 tactaaaatthaatcatttcgctatcgcgattttatatcgatctgttccatctgtcg 393  
       || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 Db    986 TATTGATTATTAATGATATCATATATGTATATGTTAATGATTGATTGTTATACGTTGTG 927  
  
 Qy    394 tgagtgtgacatcat 408  
       | | | | | | | |  
 Db    926 AATATGTTATATAAT 912

RESULT 11  
 US-08-990-571-3  
 ; Sequence 3, Application US/08990571  
 ; Patent No. 6214971  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Reed, Steven G. et al.  
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT  
 OF B. M  
 ; NUMBER OF SEQUENCES: 79  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SEED AND BERRY  
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue  
 ; CITY: Seattle  
 ; STATE: Washington  
 ; COUNTRY: USA  
 ; ZIP: 98104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/990,571  
 ; FILING DATE: 11-DEC-1997  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Maki, David J.  
 ; REGISTRATION NUMBER: 31,392  
 ; REFERENCE/DOCKET NUMBER: 210121.426C2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (206) 622-4900  
 ; TELEFAX: (206) 682-6031  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2430 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 US-08-990-571-3

Query Match 6.8%; Score 29.4; DB 4; Length 2430;  
 Best Local Similarity 51.1%; Pred. No. 6.3;  
 Matches 69; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```

Qy  274 ttatattctgctcgacaacgagtatggaattaaggcacgttatcatcagtgaatgaaacaga 333
     | ||| | | | | ||| | | | | | | | | | | | | | | | | | | | | | | |
Db  1385 TAATAAAATTAGTATACAATGATTATATTACAGATGACTATTGATTATTGTATCAATTAAA 1444

Qy  334 tactaaaatttaatcatttcgctatcgcgattttatatcgatctgtccatctgtcg 393
     | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  1445 TATTGATTATTAATGATATCATATATGTATATGTTAATGATTGATTGTTACGTTGTG 1504

Qy  394 tgagtgtgacatcat 408
     | | | | | | |
Db  1505 AATATGTTATATAAT 1519
  
```

RESULT 12  
 US-08-990-571-40/c  
 ; Sequence 40, Application US/08990571  
 ; Patent No. 6214971  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Reed, Steven G. et al.  
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT  
 OF B. M  
 ; NUMBER OF SEQUENCES: 79  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SEED AND BERRY  
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue  
 ; CITY: Seattle  
 ; STATE: Washington  
 ; COUNTRY: USA  
 ; ZIP: 98104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/990,571  
 ; FILING DATE: 11-DEC-1997  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Maki, David J.  
 ; REGISTRATION NUMBER: 31,392  
 ; REFERENCE/DOCKET NUMBER: 210121.426C2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (206) 622-4900  
 ; TELEFAX: (206) 682-6031  
 ; INFORMATION FOR SEQ ID NO: 40:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2430 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear

US-08-990-571-40

Query Match 6.8%; Score 29.4; DB 4; Length 2430;  
Best Local Similarity 51.1%; Pred. No. 6.3;  
Matches 69; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 274 ttatattctgctcgacaacgagtatggaattaaggcacgttatcagtgaatgaaacaga 333  
| || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1046 TAATAAAATTAGTATACAATGATTATATTACAGATGACTATTGATTATTGTATCAATTAAA 987

Qy 334 tactaaaatttaatcatttcgctatcgcgattttatatcgtatctgttccatctgtcg 393  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 986 TATTGATTATTAATGATATCATATATGTATATGTTAATGATTGATTGTTACGTTGTG 927

Qy 394 tgagtgtgacatcat 408  
| | | | | | | |  
Db 926 AATATGTTATATAAT 912

RESULT 13  
US-08-723-142A-3  
; Sequence 3, Application US/08723142A  
; Patent No. 6306396  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Sleath, Paul R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTIC INFECTION  
; NUMBER OF SEQUENCES: 49  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/723,142A  
; FILING DATE: 01-OCT-1996  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.426  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 2430 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-723-142A-3

Query Match 6.8%; Score 29.4; DB 4; Length 2430;  
Best Local Similarity 51.1%; Pred. No. 6.3;  
Matches 69; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
  
Qy 274 ttatattctgctcgacaacgagtatggaattaaggcacgttatcagtgaatgaaacaga 333  
| | || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1385 TAATAAAATTAGTATACAATGATTATATTACAGATGACTATTGATTATTGTATCAATTAAA 1444  
  
Qy 334 tactaaaatthaatcatttcgctatcgcgattttatatcgatctgttccatctgtcg 393  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1445 TATTGATTATTAATGATATCATATATGTATATGTTAATGATTGATTGTTACGTTGTG 1504  
  
Qy 394 tgagtgtgacatcat 408  
| | | | | | | |  
Db 1505 AATATGTTATATAAT 1519

RESULT 14  
US-08-723-142A-40/c  
; Sequence 40, Application US/08723142A  
; Patent No. 6306396  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Sleath, Paul R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; NUMBER OF SEQUENCES: 49  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/723,142A  
; FILING DATE: 01-OCT-1996  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.426  
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2430 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-723-142A-40

Query Match 6.8%; Score 29.4; DB 4; Length 2430;  
Best Local Similarity 51.1%; Pred. No. 6.3;  
Matches 69; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
  
Qy 274 ttatattctgctcgacaacgagtatggaattaaggcacgttatcatcagtgaatgaaacaga 333  
| |||| | | | |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1046 TAATAAAATTAGTATAACAATGATTATATTACAGATGACTATTGATTATTGTATCAATTAAA 987  
  
Qy 334 tactaaaatttaatcatttcgctatcgcgattttatatcgatctgttccatctgtcg 393  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 986 TATTGATTATTAATGATATCATATATGTATATGTTAATGATTGATTGTTATACGTTGTG 927  
  
Qy 394 tgagtgtgacatcat 408  
| | | | | | | |  
Db 926 AATATGTTATATAAT 912

RESULT 15  
US-08-895-601-2/c  
; Sequence 2, Application US/08895601  
; Patent No. 6060262  
; GENERAL INFORMATION:  
; APPLICANT: Beer-Romero, Peggy  
; APPLICANT: Strack, Peter J.  
; APPLICANT: Glass, Susan J.  
; APPLICANT: Rolfe, Mark  
; TITLE OF INVENTION: REGULATION OF KAPPA B (I<sub>K</sub>B) DEGRADATION,  
; TITLE OF INVENTION: AND METHODS AND REAGENTS RELATED THERETO  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-2170  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/895,601  
; FILING DATE: 16-JUL-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

; NAME: Vincent, Matthew P.  
; REGISTRATION NUMBER: 36,709  
; REFERENCE/DOCKET NUMBER: MIV-096.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-832-1000  
; TELEFAX: 617-832-7000  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2790 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 2..2782  
US-08-895-601-2

Query Match 6.8%; Score 29.4; DB 3; Length 2790;  
Best Local Similarity 60.8%; Pred. No. 6.7;  
Matches 48; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
  
Qy 310 cgttatatacgtaatgaaacagatactaaaatttaatcatttcgcgtatcgcgatttt 369  
| |||| | ||| ||| | || | || || | || |||| | || | || || |||  
Db 2524 CATTAAACAGCCTTCCAAAACCACTGTATAACCTGATGATTGCAGTGACTGTAGCCATT 2465  
  
Qy 370 atatcgatatctgttccatc 388  
| ||| | | | |||||| |||  
Db 2464 ATACTTTGTATGTTCCCTC 2446

Search completed: February 7, 2002, 11:43:03  
Job time: 9149 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 08:21:05 ; Search time 4942.22 Seconds  
(without alignments)  
945.813 Million cell updates/sec

Title: US-09-394-745-7826  
Perfect score: 435  
Sequence: 1 aattcacgggccgacgcacg.....cgtccgggtcttcctgaat 435

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

1: em\_estfun:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estom:\*

5: em\_estpl:\*

6: em\_estba:\*

7: em\_estro:\*

8: em\_estov:\*

9: em\_htc:\*

10: gb\_est1:\*

11: gb\_est2:\*

12: gb\_htc:\*

13: gb\_gss:\*

14: em\_gss\_fun:\*

15: em\_gss\_hum:\*

16: em\_gss\_inv:\*

17: em\_gss\_pln:\*

18: em\_gss\_pro:\*

19: em\_gss\_rod:\*

20: em\_gss\_vrt:\*

21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%

Result		Query						Description
	No.	Score	Match	Length	DB	ID		
c	1	186.2	42.8	883	13	BH129979		BH129979 G-6e20 Ma
c	2	89.2	20.5	327	13	BH128931		BH128931 G-4el.f M
	3	52.4	12.0	417	13	BH129483		BH129483 G-5f13.r
c	4	51.4	11.8	633	10	AI987313		AI987313 660003G09
c	5	50.6	11.6	337	13	BH130265		BH130265 G-6m9.f M
	6	47.8	11.0	585	13	BH140088		BH140088 ZMMBBb000
c	7	46.4	10.7	316	13	BH128161		BH128161 G-2m14.r
c	8	45.2	10.4	408	13	BH139925		BH139925 ZMMBBb000
	9	43.8	10.1	825	13	BH140464		BH140464 ZMMBBb000
	10	43.2	9.9	366	13	BH127373		BH127373 G-1b10.f
	11	43.2	9.9	517	13	BH128983		BH128983 G-4f7.r M
	12	42.4	9.7	839	13	BH140422		BH140422 ZMMBBb000
	13	41	9.4	850	13	BH129844		BH129844 G-6a14 Ma
	14	39.8	9.1	443	10	AW059486		AW059486 fe14f11.y
	15	39.2	9.0	550	13	AZ515621		AZ515621 BMBACR039
	16	37.8	8.7	579	13	AZ365203		AZ365203 1M0111G16
c	17	37.6	8.6	518	13	AQ844827		AQ844827 an35c05 J
c	18	37.2	8.6	1101	13	CNS000G9		AL052882 Drosophil

c	19	36.4	8.4	646	13	AZ526244	AZ526244	253PbD01
	20	36.2	8.3	420	10	AW154945	AW154945	614092E02
	21	36.2	8.3	503	10	AW163853	AW163853	614092E02
	22	36.2	8.3	605	10	AW011701	AW011701	614011H09
	23	36.2	8.3	660	10	AI783441	AI783441	614011H09
	24	36	8.3	907	13	BH128472	BH128472	G-3f5 Mai
c	25	35.8	8.2	273	11	N97589	N97589	1335C3 czap
c	26	35.8	8.2	366	13	AZ465854	AZ465854	1M0276L05
	27	35.4	8.1	441	13	AZ046475	AZ046475	nbeb0090L
c	28	35.4	8.1	626	10	AW761414	AW761414	s167b12.y
	29	35.4	8.1	820	13	AQ856532	AQ856532	nbeb0003J
c	30	35.2	8.1	247	10	AI183898	AI183898	qe23d07.x
c	31	35.2	8.1	374	10	AI001985	AI001985	ot39g06.s
c	32	35.2	8.1	386	10	AI004706	AI004706	ot95f11.x
c	33	35.2	8.1	449	10	AI083598	AI083598	ox61c09.s
c	34	35.2	8.1	597	10	AW182460	AW182460	xj42d05.x
c	35	35.2	8.1	926	13	CNS0087L	AL051525	Drosophil
	36	35.2	8.1	940	13	CNS045F1	AL275302	Tetraodon
	37	35.2	8.1	1101	13	CNS0039G	AL063921	Drosophil
	38	35	8.0	469	13	AQ535127	AQ535127	RPCI-11-3
	39	35	8.0	828	13	BH140722	BH140722	ZMMBb000
c	40	34.8	8.0	408	10	AW636289	AW636289	b145a04.w
c	41	34.8	8.0	621	13	AZ738621	AZ738621	RPCI-24-7
c	42	34.8	8.0	899	13	CNS02ZBJ	AL220744	Tetraodon
	43	34.8	8.0	1201	13	CNS016BY	AL106552	Drosophil
c	44	34.6	8.0	577	10	AI728127	AI728127	BNLGH1952
	45	34.6	8.0	693	13	AZ365021	AZ365021	1M0111G02

## ALIGNMENTS

RESULT 1  
 BH129979/c

LOCUS BH129979 883 bp DNA GSS 23-JUL-2001  
 DEFINITION G-6e20 Maize Random Small-insert Genomic Library Zea mays genomic clone G-6e20 both, DNA sequence.

ACCESSION BH129979  
 VERSION BH129979.1 GI:14998878  
 KEYWORDS GSS.  
 SOURCE Zea mays.  
 ORGANISM Zea mays  
     Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
     Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
     clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 883)  
 AUTHORS Meyers, B.C., Tingey, S.V. and Morgante, M.  
 TITLE Abundance, distribution and transcriptional activity of repetitive elements in the maize genome  
 JOURNAL Genome Res. (2001) In press  
 COMMENT Contact: Morgante M  
     Suite 200  
     Dupont Genomics  
     PO Box 6104, Newark, DE 19714-6104, USA  
     Tel: 302 631 2638  
     Fax: 302 631 2607  
     Email: Michele.morgante@usa.dupont.com

Sequences were trimmed to include only high quality bases; forward and reverse reads were assembled when significant overlaps were detected.

Seq primer: M13univ and M13reverse  
Class: shotgun.

FEATURES                    Location/Qualifiers  
source                    1. .883  
                          /organism="Zea mays"  
                          /strain="B73"  
                          /db\_xref="taxon:4577"  
                          /clone="G-6e20"  
                          /clone\_lib="Maize Random Small-insert Genomic Library"  
                          /sex="hermaphrodite"  
                          /tissue\_type="leaf"  
                          /cell\_type="Young leaf"  
                          /dev\_stage="seedling"  
                          /note="Vector: pCR-Script; Total genomic DNA was nebulized;  
                          ; ends were polished with Pfu polymerase and the fragments  
                          cloned into pCR-Script."

BASE COUNT        281 a    169 c    169 g    227 t    37 others

ORIGIN

Query Match                  42.8%; Score 186.2; DB 13; Length 883;  
Best Local Similarity      70.3%; Pred. No. 1.7e-41;  
Matches 253; Conservative 0; Mismatches 95; Indels 12; Gaps 2;

Qy    27 ggaaccttatgtgtttctggcagacatcgcccttattggtgacatctctaaattgc 86  
      ||||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db    592 GGAACCTTATGTGTTCCCTCTGGCAGATATTGTCTTATTGGTGAACATCTTAAATTGC 533

Qy    87 ttaaggcgatacatgttatgtccactagagaaaacaacatcctgagacactcaccttatt 146  
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db    532 CTACGGCGATACTTGTATGTCCACTAGAGAAACCATATCCTGAGGCACTCGTCTCGCT 473

Qy    147 tggaaatgtctcgcgattatcgctgatgtggacatgtttacatgcttctactctaa 206  
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db    472 CAGAACNNNCTTATGATTATCGCTGATANNNNCATGGTTNNNNNNNTCTCTNNNNNNAN 413

Qy    207 aagtctttgctccgaatctcgagacgagattatTTAAGGGGGAGGGCTGTAACACCC 266  
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db    412 NNGTCCTTCATTCTGNNNNCTCGGGANNNNNNTTAAGGGGGAGGGTNNTAACACCC 353

Qy    267 caggtgtttatattctgctcgacaacgagatggaaattaagcacgttatatcagtgaatg 326  
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db    352 CAGGTGT-----TCGATAATGAGTATGGATTAAAGCACGTAAAATCAGTGGATA 304

Qy    327 aaacagataactaaaattaatcat-tttcgctatcgcgatttttatatcgatctgtcc 385  
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db    303 AACCGGATGCTAAATTAAATCATCTTGCTATCGCGGTTTAATATCGCATCTGTTTC 244

RESULT    2  
BH128931/c  
LOCUS     BH128931    327 bp    DNA                    GSS        23-JUL-2001  
DEFINITION G-4el.f Maize Random Small-insert Genomic Library Zea mays genomic  
clone G-4el both, DNA sequence.

ACCESSION BH128931  
 VERSION BH128931.1 GI:14996763  
 KEYWORDS GSS.  
 SOURCE Zea mays.  
 ORGANISM Zea mays  
     Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
     Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
     clade; Panicoideae; Andropogoneae; Zea.  
 REFERENCE 1 (bases 1 to 327)  
 AUTHORS Meyers,B.C., Tingey,S.V. and Morgante,M.  
 TITLE Abundance, distribution and transcriptional activity of repetitive  
     elements in the maize genome  
 JOURNAL Genome Res. (2001) In press  
 COMMENT Contact: Morgante M  
     Suite 200  
     Dupont Genomics  
     PO Box 6104, Newark, DE 19714-6104, USA  
     Tel: 302 631 2638  
     Fax: 302 631 2607  
     Email: Michele.morgante@usa.dupont.com  
     Sequences were trimmed to include only high quality bases; forward  
     and reverse reads were assembled when significant overlaps were  
     detected.  
     Seq primer: M13univ  
     Class: shotgun.  
 FEATURES                  Location/Qualifiers  
 source                  1. .327  
                         /organism="Zea mays"  
                         /strain="B73"  
                         /db\_xref="taxon:4577"  
                         /clone="G-4el"  
                         /clone\_lib="Maize Random Small-insert Genomic Library"  
                         /sex="hermaphrodite"  
                         /tissue\_type="leaf"  
                         /cell\_type="Young leaf"  
                         /dev\_stage="seedling"  
                         /note="Vector: pCR-Script; Total genomic DNA was nebulized  
                         ; ends were polished with Pfu polymerase and the fragments  
                         cloned into pCR-Script."  
 BASE COUNT        102 a     61 c.    74 g     88 t     2 others  
 ORIGIN

Query Match                  20.5%; Score 89.2; DB 13; Length 327;  
 Best Local Similarity    87.1%; Pred. No. 1.6e-14;  
 Matches 108; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

Qy	27	ggaaccttatgtttctggcagac-atcgccctattggacatctctaaattag	85
Db	124	GGAACCTTATGTGTTCTCTGGCGAACAAATCGCCTATTGGACATCTCAAATTAG	65
Qy	86	cttaaggcgatacatgttatgtccactagagaaacaacatcctgagacactcacttat	145
Db	64	NNTAAGACGATACATGTTCTGTCCACAAGAGAACATCTTGAGACACTTCTCGC	5
Qy	146	ttgg 149	

Db 4 TTGG 1

RESULT 3  
 BH129483  
 LOCUS BH129483 417 bp DNA GSS 23-JUL-2001  
 DEFINITION G-5f13.r Maize Random Small-insert Genomic Library Zea mays genomic  
 clone G-5f13 both, DNA sequence.  
 ACCESSION BH129483  
 VERSION BH129483.1 GI:14997879  
 KEYWORDS GSS.  
 SOURCE Zea mays.  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Zea.  
 REFERENCE 1 (bases 1 to 417)  
 AUTHORS Meyers, B.C., Tingey, S.V. and Morgante, M.  
 TITLE Abundance, distribution and transcriptional activity of repetitive  
 elements in the maize genome  
 JOURNAL Genome Res. (2001) In press  
 COMMENT Contact: Morgante M  
 Suite 200  
 Dupont Genomics  
 PO Box 6104, Newark, DE 19714-6104, USA  
 Tel: 302 631 2638  
 Fax: 302 631 2607  
 Email: Michele.morgante@usa.dupont.com  
 Sequences were trimmed to include only high quality bases; forward  
 and reverse reads were assembled when significant overlaps were  
 detected.  
 Seq primer: M13reverse  
 Class: shotgun.  
 FEATURES Location/Qualifiers  
 source 1. .417  
 /organism="Zea mays"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="G-5f13"  
 /clone\_lib="Maize Random Small-insert Genomic Library"  
 /sex="hermaphrodite"  
 /tissue\_type="leaf"  
 /cell\_type="Young leaf"  
 /dev\_stage="seedling"  
 /note="Vector: pPCR-Script; Total genomic DNA was nebulized  
 ; ends were polished with Pfu polymerase and the fragments  
 cloned into pPCR-Script."  
 BASE COUNT 98 a 95 c 75 g 145 t 4 others  
 ORIGIN

Query Match 12.0%; Score 52.4; DB 13; Length 417;  
 Best Local Similarity 69.4%; Pred. No. 0.0003;  
 Matches 100; Conservative 0; Mismatches 41; Indels 3; Gaps 2;

Qy 182 gtgttacatgttctactcttaaaagtctttgctccgaatctcgagacgagatt-at 240

```

Db      99 GAGTTACATGCTTCTCCACCCTTAAA--TATCCTCATTGAATCTCGGGACGAGATTCTT 156
Qy     241 tttaaggggggagggctgtaacaccccccaggtttatattctgctcgacaacgagtatgg 300
          ||||||| ||||| ||||||| ||||||| ||||| | | | | | | |
Db     157 TTTAAGGGGGGAAGGCTGTGACACCCCCAGGTGTCTATTCGCCTTATATCGGGAGATTAA 216
Qy     301 aattaagcacgttatatcagtcaa 324
          || | | | | | | |
Db     217 TCCCAATCTCGGATGCTCAGTAAA 240

```

```

RESULT      4
AI987313/c
LOCUS      AI987313      633 bp      mRNA           EST      01-SEP-1999
DEFINITION 660003G09.x1 660 - Mixed stages of anther and pollen Zea mays cDNA,
mRNA sequence.
ACCESSION   AI987313
VERSION     AI987313.1  GI:5816397
KEYWORDS    EST.
SOURCE      Zea mays.
ORGANISM   Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 633)
AUTHORS    Walbot,V.
TITLE      Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL    Unpublished (1999)
COMMENT    Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 660003 row: G column: 09.
FEATURES    Location/Qualifiers
source      1. .633
            /organism="Zea mays"
            /cultivar="Ohio43"
            /db_xref="taxon:4577"
            /clone_lib="660 - Mixed stages of anther and pollen"
            /tissue_type="whole premeiotic anthers to pollen shed"
            /dev_stage="premeiotic anthers to pollen shed"
            /lab_host="XLOLR"
            /note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
Site_2: XhoI; Anther and pollen cDNA library.
Directionally sequenced with 5' end at the EcoRI site.
Created by Amie Franklin."
BASE COUNT  145 a    194 c    166 g    128 t
ORIGIN
```

```

Query Match      11.8%;  Score 51.4;  DB 10;  Length 633;
Best Local Similarity 90.2%;  Pred. No. 0.00061;
Matches 55;  Conservative 0;  Mismatches 6;  Indels 0;  Gaps 0;
```

Qy 265 cccaggtttatattctgctcgacaacgagtatggattaaggcacgttatcatgtaa 324  
 ||||| ||||||||||||||||||||||||||||||| ||||||| | | |||||||||  
 Db 61 CCCATGTGTTATATTCTGCTCGACAAACGAGTATGGATTAAAGCGATAATATCAGTGG 2

Qy 325 t 325  
 |  
 Db 1 T 1

**RESULT** 5  
**BH130265/c**  
**LOCUS** BH130265 337 bp DNA GSS 23-JUL-2001  
**DEFINITION** G-6m9.f Maize Random Small-insert Genomic Library Zea mays genomic clone G-6m9 both, DNA sequence.  
**ACCESSION** BH130265  
**VERSION** BH130265.1 GI:14999460  
**KEYWORDS** GSS.  
**SOURCE** Zea mays.  
**ORGANISM** Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Zea.  
**REFERENCE** 1 (bases 1 to 337)  
**AUTHORS** Meyers,B.C., Tingey,S.V. and Morgante,M.  
**TITLE** Abundance, distribution and transcriptional activity of repetitive elements in the maize genome  
**JOURNAL** Genome Res. (2001) In press  
**COMMENT** Contact: Morgante M  
 Suite 200  
 Dupont Genomics  
 PO Box 6104, Newark, DE 19714-6104, USA  
 Tel: 302 631 2638  
 Fax: 302 631 2607  
 Email: Michele.morgante@usa.dupont.com  
 Sequences were trimmed to include only high quality bases; forward and reverse reads were assembled when significant overlaps were detected.  
 Seq primer: M13univ  
 Class: shotgun.  
**FEATURES** Location/Qualifiers  
**source** 1. .337  
 /organism="Zea mays"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="G-6m9"  
 /clone\_lib="Maize Random Small-insert Genomic Library"  
 /sex="hermaphrodite"  
 /tissue\_type="leaf"  
 /cell\_type="Young leaf"  
 /dev\_stage="seedling"  
 /note="Vector: pCR-Script; Total genomic DNA was nebulized ; ends were polished with Pfu polymerase and the fragments cloned into pCR-Script."

**BASE COUNT** 76 a 77 c 69 g 111 t 4 others  
**ORIGIN**

Query Match 11.6%; Score 50.6; DB 13; Length 337;  
 Best Local Similarity 76.5%; Pred. No. 0.00094;  
 Matches 62; Conservative 0; Mismatches 19; Indels 0; Gaps 0;  
  
 Qy 199 actcttaaaagtctttgctccgaatctcgagacgagattatTTAAGGGGGAGGGCTG 258  
 || || || | | | | || || || || || || || || || || || || || || || || || ||  
 Db 217 ACCCTCCAAGGGACTCTACCAAAATCTCGGGACGAGATTCTTAAGGGGGAGGGCTG 158  
  
 Qy 259 taacaccccaggtgttatat 279  
 || || || || || || || || || || || || || || || || || || || || || || || ||  
 Db 157 TAACACCCCCAGGTGTTACCAT 137

RESULT 6  
 BH140088  
 LOCUS BH140088 585 bp DNA GSS 07-AUG-2001  
 DEFINITION ZMMBBb0001H02f Maize B73 Zea mays genomic clone ZMMBBb0001H02f, DNA sequence.  
 ACCESSION BH140088  
 VERSION BH140088.1 GI:15099149  
 KEYWORDS GSS.  
 SOURCE Zea mays.  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Zea.  
 REFERENCE 1 (bases 1 to 585)  
 AUTHORS Tomkins, J.P., Main, D., Goicoechea, J.L., Frisch, D.A. and Wing, R.A.  
 TITLE A Deep-Coverage BAC Library for Maize  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu  
 Seq primer: TAATACGACTCACTATAGGG  
 Class: BAC ends  
 High quality sequence stop: 584.  
 FEATURES Location/Qualifiers  
 source 1. .585  
 /organism="Zea mays"  
 /strain="B73"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBBb0001H02f"  
 /clone\_lib="Maize B73"  
 /tissue\_type="Young leaves"  
 /lab\_host="E. coli"  
 /note="Vector: pCUGIBAC-1; Site\_1: HindIII; Site\_2: NotI;  
 For more details on library preparation, ordering clones  
 and sequence analysis see  
<http://www.genome.clemson.edu/projects/stc/maize/ZMMBBb>"  
 BASE COUNT 180 a 143 c 128 g 134 t  
 ORIGIN

```

Query Match           11.0%;  Score 47.8;  DB 13;  Length 585;
Best Local Similarity 77.3%;  Pred. No. 0.0061;
Matches   58;  Conservative   0;  Mismatches   17;  Indels   0;  Gaps   0;

Qy      200 ctcttaaaagtctttgctccgaatctcgagacgagattatTTAAGGGGGAGGGCTGT 259
        | || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      508 CCCTCCAAGGGACTCTACCTAAAATCTCGGGACGATATTCTTAAGGGGGAGGGCTGT 567

Qy      260 aacaccccaggtgtt 274
        | | | | | | | | | |
Db      568 AACACCCCAGGTGTT 582

```

RESULT 7  
BH128161/c  
LOCUS BH128161 316 bp DNA GSS 23-JUL-2001  
DEFINITION G-2m14.r Maize Random Small-insert Genomic Library Zea mays genomic  
clone G-2m14 both, DNA sequence.  
ACCESSION BH128161  
VERSION BH128161.1 GI:14995993  
KEYWORDS GSS.  
SOURCE Zea mays.  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 316)  
AUTHORS Meyers, B.C., Tingey, S.V. and Morgante, M.  
TITLE Abundance, distribution and transcriptional activity of repetitive  
elements in the maize genome  
JOURNAL Genome Res. (2001) In press  
COMMENT Contact: Morgante M  
Suite 200  
Dupont Genomics  
PO Box 6104, Newark, DE 19714-6104, USA  
Tel: 302 631 2638  
Fax: 302 631 2607  
Email: Michele.morgante@usa.dupont.com  
Sequences were trimmed to include only high quality bases; forward  
and reverse reads were assembled when significant overlaps were  
detected.  
Seq primer: M13reverse  
Class: shotgun.  
FEATURES Location/Qualifiers  
source 1. .316  
/organism="Zea mays"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="G-2m14"  
/clone\_lib="Maize Random Small-insert Genomic Library"  
/sex="hermaphrodite"  
/tissue\_type="leaf"  
/cell\_type="Young leaf"  
/dev\_stage="seedling"  
/note="Vector: pCR-Script; Total genomic DNA was nebulized"

; ends were polished with Pfu polymerase and the fragments cloned into pCR-Script."

BASE COUNT      70 a      66 c      70 g      107 t      3 others  
ORIGIN

Query Match      10.7%; Score 46.4; DB 13; Length 316;  
Best Local Similarity 89.3%; Pred. No. 0.014;  
Matches 50; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy    219 ccgaatctcgagacgagattatttaaggggggagggtgtaacaccccaggttt 274  
      || ||||| ||||||||| ||||||||||| ||||||||||| ||||||||| |||||  
Db    196 CCCAATCTCAGGACGAGATTCTTAAGGGGGAGGGCTGTAACACCCCTGGTGT 141

RESULT 8  
BH139925/c  
LOCUS    BH139925      408 bp      DNA      GSS      07-AUG-2001  
DEFINITION ZMMBBb0001A14f Maize B73 Zea mays genomic clone ZMMBBb0001A14f, DNA sequence.  
ACCESSION BH139925  
VERSION    BH139925.1 GI:15098986  
KEYWORDS GSS.  
SOURCE    Zea mays.  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 408)  
AUTHORS Tomkins, J.P., Main, D., Goicoechea, J.L., Frisch, D.A. and Wing, R.A.  
TITLE A Deep-Coverage BAC Library for Maize  
JOURNAL Unpublished (2001)  
COMMENT Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: TAATACGACTCACTATAGGG  
Class: BAC ends  
High quality sequence stop: 405.  
FEATURES      Location/Qualifiers  
source      1..408  
/organism="Zea mays"  
/strain="B73"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBBb0001A14f"  
/clone\_lib="Maize B73"  
/tissue\_type="Young leaves"  
/lab\_host="E. coli"  
/note="Vector: pCUGIBAC-1; Site\_1: HindIII; Site\_2: NotI;  
For more details on library preparation, ordering clones  
and sequence analysis see  
<http://www.genome.clemson.edu/projects/stc/maize/ZMMBBb>"  
BASE COUNT      118 a      81 c      72 g      137 t

ORIGIN

Query Match 10.4%; Score 45.2; DB 13; Length 408;  
Best Local Similarity 86.2%; Pred. No. 0.031;  
Matches 50; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 222 aatctcgagacgagattatTTaaGGGGGAGGGCTGtaacACCCcaggTtttat 279  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 286 AATCTCGGGACGAGATTCTTTATGGGGGAAGGATGTAACACCCCTGGTGTACTAT 229

RESULT 9  
BH140464  
LOCUS BH140464 825 bp DNA GSS 07-AUG-2001  
DEFINITION ZMMBb0002F03r Maize B73 Zea mays genomic clone ZMMBb0002F03r, DNA sequence.  
ACCESSION BH140464  
VERSION BH140464.1 GI:15099525  
KEYWORDS GSS.  
SOURCE Zea mays.  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 825)  
AUTHORS Tomkins, J.P., Main, D., Goicoechea, J.L., Frisch, D.A. and Wing, R.A.  
TITLE A Deep-Coverage BAC Library for Maize  
JOURNAL Unpublished (2001)  
COMMENT Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Class: BAC ends  
High quality sequence start: 56  
High quality sequence stop: 788.  
FEATURES Location/Qualifiers  
source 1. .825  
/organism="Zea mays"  
/strain="B73"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBb0002F03r"  
/clone\_lib="Maize B73"  
/tissue\_type="Young leaves"  
/lab\_host="E. coli"  
/note="Vector: pCUGIBAC-1; Site\_1: HindIII; Site\_2: NotI;  
For more details on library preparation, ordering clones  
and sequence analysis see  
<http://www.genome.clemson.edu/projects/stc/maize/ZMMBb>"  
BASE COUNT 259 a 160 c 166 g 236 t 4 others  
ORIGIN

**RESULT** 10  
**BH127373**  
**LOCUS** BH127373 366 bp DNA GSS 23-JUL-2001  
**DEFINITION** G-1b10.f Maize Random Small-insert Genomic Library Zea mays genomic clone G-1b10 both, DNA sequence.  
**ACCESSION** BH127373  
**VERSION** BH127373.1 GI:14995205  
**KEYWORDS** GSS.  
**SOURCE** Zea mays.  
**ORGANISM** Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
**REFERENCE** 1 (bases 1 to 366)  
**AUTHORS** Meyers, B.C., Tingey, S.V. and Morgante, M.  
**TITLE** Abundance, distribution and transcriptional activity of repetitive elements in the maize genome  
**JOURNAL** Genome Res. (2001) In press  
**COMMENT** Contact: Morgante M  
Suite 200  
Dupont Genomics  
PO Box 6104, Newark, DE 19714-6104, USA  
Tel: 302 631 2638  
Fax: 302 631 2607  
Email: Michele.morgante@usa.dupont.com  
Sequences were trimmed to include only high quality bases; forward and reverse reads were assembled when significant overlaps were detected.  
Seq primer: M13univ  
Class: shotgun.  
**FEATURES**  
**source** Location/Qualifiers  
1. .366  
/organism="Zea mays"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="G-1b10"  
/clone\_lib="Maize Random Small-insert Genomic Library"  
/sex="hermaphrodite"  
/tissue\_type="leaf"  
/cell\_type="Young leaf"  
/dev\_stage="seedling"  
/note="Vector: pCR-Script; Total genomic DNA was nebulized ; ends were polished with Pfu polymerase and the fragments cloned into pCR-Script."

BASE COUNT        117 a        83 c        76 g        83 t        7 others  
ORIGIN

Query Match                    9.9%;    Score 43.2;   DB 13;   Length 366;  
Best Local Similarity    75.0%;   Pred. No. 0.11;  
Matches    54;   Conservative    0;   Mismatches    18;   Indels    0;   Gaps    0;  
  
Qy    202 cttaaaaagtctttgctccgaatctcgagacgagattttaaaggggggagggtgtaa 261  
      |||    |||    |    |    |||||    |||    |    |||||    |||    |||||  
Db    295 CTCTGAAGAATCCCGACTCGAATTTCGGGGCGAGATTCTTTAAGAGGGTAGGGCTGTAA 354  
  
Qy    262 caccccaggtgt 273  
      |||||    |||||  
Db    355 CACCCTAGGTGT 366

RESULT 11  
BH128983  
LOCUS      BH128983      517 bp      DNA      GSS      23-JUL-2001  
DEFINITION G-4f7.r Maize Random Small-insert Genomic Library Zea mays genomic  
clone G-4f7 both, DNA sequence.  
ACCESSION BH128983  
VERSION     BH128983.1    GI:14996828  
KEYWORDS    GSS.  
SOURCE      Zea mays.  
ORGANISM    Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 517)  
AUTHORS Meyers,B.C., Tingey,S.V. and Morgante,M.  
TITLE      Abundance, distribution and transcriptional activity of repetitive  
elements in the maize genome  
JOURNAL    Genome Res. (2001) In press  
COMMENT    Contact: Morgante M  
Suite 200  
Dupont Genomics  
PO Box 6104, Newark, DE 19714-6104, USA  
Tel: 302 631 2638  
Fax: 302 631 2607  
Email: Michele.morgante@usa.dupont.com  
Sequences were trimmed to include only high quality bases; forward  
and reverse reads were assembled when significant overlaps were  
detected.  
Seq primer: M13reverse  
Class: shotgun.  
FEATURES      Location/Qualifiers  
source        1..517  
              /organism="Zea mays"  
              /strain="B73"  
              /db\_xref="taxon:4577"  
              /clone="G-4f7"  
              /clone\_lib="Maize Random Small-insert Genomic Library"  
              /sex="hermaphrodite"  
              /tissue\_type="leaf"  
              /cell\_type="Young leaf"



```

/db_xref="taxon:4577"
/clone="ZMMBBb0002D13r"
/clone_lib="Maize B73"
/tissue_type="Young leaves"
/lab_host="E. coli"
/note="Vector: pCUGIBAC-1; Site_1: HindIII; Site_2: NotI;
For more details on library preparation, ordering clones
and sequence analysis see
http://www.genome.clemson.edu/projects/stc/maize/ZMMBBb "
BASE COUNT      258 a      162 c      176 g      239 t      4 others
ORIGIN .

```

```

Query Match          9.7%;  Score 42.4;  DB 13;  Length 839;
Best Local Similarity 69.0%;  Pred. No. 0.2;
Matches 58;  Conservative 0;  Mismatches 26;  Indels 0;  Gaps 0;
Qy    193 ttctctactctaaaagtctttgctccgaatctcgagacgagatttttaagggggga 252
      ||||| ||||| ||||| ||| | | | ||||||| ||||||| ||| ||| ||| ||| |
Db    515 TCCTTACCAATTACCTACCCTAGGATTTAATCTCGGGACGAGATTCTTATGGGGGA 574
Qy    253 gggctgttaaacaccccaggtgtta 276
      ||| ||||||||| | | |||
Db    575 AGGATGTAAACACCCCTGGTGTAA 598

```

```

RESULT 13
BH129844
LOCUS     BH129844      850 bp      DNA          GSS      23-JUL-2001
DEFINITION G-6a14 Maize Random Small-insert Genomic Library Zea mays genomic
clone G-6a14 both, DNA sequence.
ACCESSION  BH129844
VERSION    BH129844.1  GI:14998606
KEYWORDS   GSS.
SOURCE     Zea mays.
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 850)
AUTHORS    Meyers,B.C., Tingey,S.V. and Morgante,M.
TITLE      Abundance, distribution and transcriptional activity of repetitive
elements in the maize genome
JOURNAL   Genome Res. (2001) In press
COMMENT    Contact: Morgante M
            Suite 200
            Dupont Genomics
            PO Box 6104, Newark, DE 19714-6104, USA
            Tel: 302 631 2638
            Fax: 302 631 2607
            Email: Michele.morgante@usa.dupont.com
            Sequences were trimmed to include only high quality bases; forward
            and reverse reads were assembled when significant overlaps were
            detected.
            Seq primer: M13univ and M13reverse
            Class: shotgun.
FEATURES   Location/Qualifiers

```

source 1. .850  
 /organism="Zea mays"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="G-6a14"  
 /clone\_lib="Maize Random Small-insert Genomic Library"  
 /sex="hermaphrodite"  
 /tissue\_type="leaf"  
 /cell\_type="Young leaf"  
 /dev\_stage="seedling"  
 /note="Vector: pCR-Script; Total genomic DNA was nebulized;  
 ; ends were polished with Pfu polymerase and the fragments  
 cloned into pCR-Script."  
 BASE COUNT 284 a 144 c 172 g 242 t 8 others  
 ORIGIN

Query Match 9.4%; Score 41; DB 13; Length 850;  
 Best Local Similarity 82.5%; Pred. No. 0.5;  
 Matches 47; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 222 aatctcgagacgagattatttaaggggggggctgtacaccccccagggtttata 278  
 ||||||| ||||||||| ||||||| ||||||| ||||||| ||| ||||||||| | ||| |||  
 Db 31 AATCTCGGGACGAGATTCTTTATGGGGGGAGGATGTAACACCCCTAGCGTTACTA 87

RESULT 14  
 AW059486  
 LOCUS AW059486 443 bp mRNA EST 07-JUN-2001  
 DEFINITION fe14f11.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA clone  
 IMAGE:3738861 5', mRNA sequence.  
 ACCESSION AW059486  
 VERSION AW059486.1 GI:5935125  
 KEYWORDS EST.  
 SOURCE zebrafish.  
 ORGANISM Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Rasborinae; Danio.  
 REFERENCE 1 (bases 1 to 443)  
 AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy  
 ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood  
 ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,  
 Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,  
 Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.  
 and Wilson,R.  
 TITLE WashU Zebrafish EST Project 1998  
 JOURNAL Unpublished (1998)  
 COMMENT Other\_ESTs: fe14f11.x1  
 Contact: Stephen L. Johnson  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: zebrafish@watson.wustl.edu  
 cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:  
 Matthew Clark. DNA Sequencing by: Washington University Genome

Sequencing Center Clone distribution: Genome Systems, St. Louis, Missouri (web address: [www.genomesystems.com](http://www.genomesystems.com)) (email contact: [info@genomesystems.com](mailto:info@genomesystems.com)) and Research Genetics, Huntsville, Alabama (web address: [www.resgen.com](http://www.resgen.com)) (email contact: [info@resgen.com](mailto:info@resgen.com)) and RessourcenZentrumPrimärDatenbank, Berlin, Germany (web address: [www.rzpd.de](http://www.rzpd.de))

Seq primer: T3 ET from Amersham.

FEATURES Location/Qualifiers  
source 1. .443  
/organism="Danio rerio"  
/db\_xref="taxon:7955"  
/clone="IMAGE:3738861"  
/clone\_lib="Zebrafish WashU MPIMG EST"  
/sex="mixed"  
/tissue\_type="26 somite embryos, adult livers, shield stage embryos"  
/lab\_host="XL1-blue MRF"  
/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; 1st strand cDNA was primed with a Not I - oligo(dT)15 primer [5'pGACTAGTTCTAGATCGCGAGCGGCCCTTTTTTTTT3']; double-stranded cDNA was ligated to Sal I adaptors (BRL), digested with Not I and cloned into the Not I and Sal I sites of the pSPORT1 vector (BRL). Library was constructed by Matthew Clark (Lehrach lab; ICRF, London and Max Planck Institut fuer Molekulare Genetik, Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

BASE COUNT 140 a 74 c 68 g 161 t  
ORIGIN

```

Query Match          9.1%;  Score 39.8;  DB 10;  Length 443;
Best Local Similarity 54.4%;  Pred. No. 1;
Matches 80;  Conservative 0;  Mismatches 67;  Indels 0;  Gaps 0;

Qy      273 tttatattctgctcgacaacgagtatgaaattaaggcacgttatatcagtgaatgaaacag 332
        ||||| ||| | | | ||| ||| ||| ||| | | | | | | | | | | | | | | | |
Db      176 TTTTTATAATGATTATAATCAGTTGGCATTCACACAGTTGTTATTCGATTAAATCAC 235

Qy      333 atactaaaatthaatcatttcgcatacgcgattttatatcgatatctgttccatctgtc 392
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      236 ACATCACATATTATTATTGTGTTTGTGTTTATATAATTCTTATTCCATTGGTT 295

Qy      393 gtgagtgtgacatcattttattcgtc 419
        | | | | | | | | | | | | | | | |
Db      296 GAAAGTTCATATCATTGTAATTGTC 322

```

RESULT 15

**AZ515621**  
**LOCUS** AZ515621 550 bp DNA GSS 05-OCT-2000  
**DEFINITION** BMBACR039SP6 Brugia malayi Genomic Bac Library 1 & 2 Brugia malayi genomic, DNA sequence.  
**ACCESSION** AZ515621  
**VERSION** AZ515621.1 GI:10696940  
**KEYWORDS** GSS.  
**SOURCE** Brugia malayi.  
**ORGANISM** Brugia malayi  
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
Onchocercidae; Brugia.  
**REFERENCE** 1 (bases 1 to 550)  
**AUTHORS** Daub,J., Ware,J., Foster,J., Giuliano,D., Slatko,B. and Blaxter,M.  
**TITLE** Genome survey sequences from the human parasitic nematode Brugia malayi  
**JOURNAL** Unpublished (2000)  
**COMMENT** Contact: Blaxter ML  
Institute of Cell, Animal and Population Biology  
University of Edinburgh  
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK  
Tel: +44 131 650 6760  
Fax: +44 131 670 5450  
Email: mark.blaxter@ed.ac.uk  
Sequenced from the Filarial Genome Project's Brugia malayi BAC library constructed by Jesse Pope-Chappel and Jeremy Foster. The sequence was generated by Barton Slatko, New England Biolabs, 32 Tozer Road, Beverley, MA, 01915-55110, USA.  
Seq primer: SP6 (CGCCAAGCTATTTAGGTGACAC)  
Class: BAC ends.  
**FEATURES** Location/Qualifiers  
**source** 1. .550  
/organism="Brugia malayi"  
/strain="TRS"  
/db\_xref="taxon:6279"  
/clone\_lib="Brugia malayi Genomic Bac Library 1 & 2"  
/sex="Mixed (male and female)"  
/tissue\_type="whole parasite"  
/dev\_stage="adult"  
/note="Vector: pBeloBAC III; Site\_1: Hind III; Brugia malayi genomic DNA was partially cleaved with Hind III and size fractionated. 18,000 clones were generated from 2 libraries with mean insert size 60 kbp. The library was constructed by Jesse Pope-Chappel, Smith College Northhampton MA and Dr Jeremy Foster, New England Biolabs, MA."  
**BASE COUNT** 184 a 78 c 66 g 195 t 27 others  
**ORIGIN**

Query Match 9.0%; Score 39.2; DB 13; Length 550;  
Best Local Similarity 50.6%; Pred. No. 1.5;  
Matches 89; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 71 acatctctaaattagcttaaggcgatacatgttatgtccactagagaaaacaacatcctga 130  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 341 ACAAAATATCACTGAGAATATTGAAATAANTGTTTCTGCTGCTGATAGAAATATCCTCT 400

Qy 131 gacactcaccttataatggaaatgtctcgcgattatcgctgatgtggacatgtgttacat 190  
| || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 401 AAAACGTACTATTCTTNGCCACATTCACTTGATGGTGAATAAAACTTATTGTNTCAT 460

Qy 191 gcttctctactctaaaagtctttgctccgaatctcgagacgagattatTTtaag 246  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 461 AAATAANTTAGTGCAAAGTTACTGCTATGTGACTAGATGATATAAATTAAAG 516

Search completed: February 7, 2002, 08:21:08  
Job time: 18145 sec